Real-Time Plasmid Monitoring of Batch and Fed-Batch *Escherichia coli* Cultures by NIR Spectroscopy

Marta B. Lopes[1,2,3], Kevin C. Sales[1], Vitor V. Lopes[2], Cecília R. C. Calado[1,ξ]

[1]Engineering Faculty, Catholic University of Portugal, Rio de Mouro, Portugal
[2]National Laboratory for Energy and Geology, Lisbon, Portugal
[3]Instituto de Telecomunicações, Instituto Superior Técnico, Lisboa, Portugal
[ξ] c.calado@fe.lisboa.ucp.pt

Abstract —**The development of *in-situ* monitoring techniques enabling the real-time acquisition of information concerning the key variables over different *Escherichia coli* cultivation conditions and strategies is a crucial step towards the optimization of a plasmid bioproduction process. This work shows the use of a Near-InfraRed (NIR) fiber optic probe immersed in the culture broth for the real-time acquisition of NIR spectra along different *E. coli* cultures conducted with mixtures of the carbon sources glucose and glycerol, and performed in both batch and fed-batch modes. Accurate partial least squares models based on the acquired spectral data over such different cultivation conditions were built, yielding a $R^2 \geq 0.97$ for biomass and plasmid productions and a RMSEP of 0.34 and 7.52, respectively; a R^2 of 0.93 and a RMSEP of 0.46 and 0.33 was obtained for glucose and glycerol, respectively; the acetate model produced a R^2 of 0.96 and a RMSEP of 0.32.**

Plasmid; Monitoring; Control; Fourier transform infrared; Near-infrared; spectroscopy; Escherichia coli

I. INTRODUCTION

As plasmids are increasingly being used as alternative to viral vectors, mainly due to security issues, it is urgent to develop economic plasmid production processes [1]. Plasmids are generally produced using *Escherichia coli* as the cell host. To achieve reproducible culture processes, and given the natural variability associated to processes using living cells, the development of *in-situ* monitoring techniques enabling the acquisition of information concerning the bioprocess key variables in real time is highly desired. Such variables are the bacteria growth, the plasmid production, the carbon-sources consumption (usually glucose and glycerol) and the production and consumption of the by-product acetate. Real-time information about these variables would allow the implementation of control methodologies ensuring consistency of the final product, in accordance to Process Analytical Technology (PAT) guidance, recommended by the Food and Drug Administration [2] and the European Medicines Agency [3]. The techniques developed should be applicable to the highly dynamic bioreactor environment, which due to the high specific growth rates on glucose or glycerol may present a large variation in stirring speed (varying up to 1000 rpm of Rushton Turbines) and air flow rates up to 1.5 vvm. Moreover, in order to speed up the optimization procedure of a particular plasmid bioprocess, the *in-situ* monitoring technique should also be applicable to a wide range of culture conditions. It is usual in optimization protocols the evaluation of the effect of different media compositions, such as the carbon sources glucose and glycerol, and the batch and fed-batch cultivation strategies [4].

The use of Mid-InfraRed (MIR) spectroscopy for a rapid (*at-line*) characterization of a plasmid bioproduction process has already been demonstrated by the present authors [5]. However, for this analysis samples need to be extracted from the bioreactor and subsequently dehydrated before spectra acquisition, which increases the risk of bioreactor contamination and inputs a time delay in the bioprocess characterization. Near-InfraRed (NIR) spectroscopy presents the great advantage over its neighbor MIR region for *on-line* data acquisition directly from the aqueous culture broth, due to the lower interference by water. Previous studies have shown the potential of using a NIR fiber optic probe for *in-situ* monitoring the biomass, nutrients, products and by-products from bacterial cultures [6-10].

This work evaluates the *in-situ* monitoring of the plasmid bioproduction process, over high cell densities of *E. coli* cultures, by using a NIR transflectance fiber optic probe. Multiple cultures, representing batch and fed-batch cultivation strategies, and using mixtures of the carbon sources glucose and glycerol, were combined for developing partial least squares (PLS) regression models to predict in real time the key variables of the bioprocess: biomass and plasmid production, glucose and glycerol consumption, and the production and consumption of the by-product acetate. The models developed, by accounting for a wide range of culture conditions, will be very helpful for speeding the optimization of the plasmid bioproduction process.

II. MATERIALS AND METHODS

Cultivation

Escherichia coli DH5-α containing the plasmid model pVAX-LacZ (Invitrogen, USA) was used. The cultivation was performed in a 2 l bioreactor (Biostat MD, B. Braun, Germany) with a 1.8 l working volume, in absence of antibiotic. Cultivation was maintained at pH 7.0 by automatic control through 1 M NaOH (Fluka, Switzerland) or 1 M HCl

(Sigma-Aldrich, Germany) addition, and at 37 °C with a minimal DOC of 30 % of air saturation in general by automatic adjustment of the agitation rate. An air flow rate range between 1.0 and 1.5 vvm (volume of air/volume of medium/minute) was used. The initial batch cultivation media of the three cultures studied contained 10 g/l of yeast extract (Difco, USA), 20 g/l bactotryptone (BD, UK) and 6 g/l of glycerol (cultivation A), 6 g/l of glucose (cultivation B), and 6 g/l of glycerol and 6 g/l of glucose (cultivation C). An exponential feeding phase after the consumption of the acetate produced during the batch phase was started on cultivations B and C, considering a constant specific growth rate of 0.18 1/h, a biomass yield on glucose of 0.6 and a 0.3 l feeding medium containing 45 g glucose, 22.5 g yeast extract and 22.5 g bactotryptone. Samples were taken from the bioreactor along the culture, and subsequently used for *off-line* analysis of biomass, glucose, glycerol, acetate and plasmid.

Reference analyses

Biomass in units of dry cell weight per volume of culture medium (g/l) was determined by centrifuging the cultivation samples, washing the pellet with 0.9 % (w/v) sodium chloride and drying at 80°C until constant weight. The bacterial cell pellet and the supernatant obtained from sample centrifugation (Hermle Z160M, Germany) were frozen at -20°C. Glucose, glycerol and acetate were determined by HPLC with a L-6200 Intelligent Pump (Merck-Hitachi, UK), a L-7490 LaCrom-Ri-detector (Merck, Germany), a D-2500 Chromato-integrator (Merck-Hitachi, Germany) and an Aminex® Fermentation Monitor HPLC column for culture broth monitoring (Bio-Rad, USA) maintained at 50 °C, and by using H_2SO_4 at 0.6 ml/min as eluent. Plasmids were extracted from the bacteria cell by the alkaline cell lysis method, and subsequent plasmid concentration and purity degree were determined by hydrophobic interaction HPLC, as described in [5].

NIR spectroscopy and chemometric analysis

NIR spectra were obtained using an NIR transflection fiber optic probe IN-271P (Bruker Optics, Germany), with a pathlength of 2 mm, coupled to a Vertex-70 spectrometer (Bruker Optics, Germany) with a TE-InGaAs detector. The fiber optic probe was submerged in the bioreactor and stem sterilized simultaneously with the cultivation medium. NIR spectra were collected every 2 minutes in the 12500-5400 cm^{-1} (800-1851 nm) range, consisting of 32 coadded scans with 8 cm^{-1} resolution (2 nm steps). The scanner velocity was set to 20 kHz and the aperture setting defined was 6 mm.

PLS models were built for biomass, glucose, glycerol, acetate and plasmid concentrations. The predictive performance of each PLS model was assessed by the coefficient of determination (R^2) and the root mean squared error of prediction (RMSEP). Both values were calculated based on a test set when there is a large number of samples, and on leave-one-out (LOO) cross-validation for smaller datasets.

The following spectral pre-processing techniques were explored to improve the predictive performance of the models: constant offset elimination, min-max normalization, standard normal variate (SNV), multiplicative scatter correction (MSC) and spectral derivatives. Wavelength selection was also performed to discard the spectral regions with irrelevant information with the variables under study, by dividing the spectral window into 10 sub-regions. The best combination of spectral regions and the pre-processing techniques was selected by picking the PLS model with the smallest RMSEP.

III. Results and Discussion

Plasmid was first produced in an *E. coli* culture conducted in non-selective and complex media with 6g/l glycerol as carbon source (cultivation A), resulting in high plasmid yields per biomass, of 6.0 mg/g, compared to others [11,12]. For that culture, and for each analyte, several PLS regression models were obtained by combining several pre-processing techniques. For the biomass and plasmid models, the RMSEP was calculated based on an independent test validation dataset, given the larger number of samples available. For the glycerol and acetate models the RMSEP was obtained by leave-one-out (LOO) cross-validation, as fewer samples were available, provided by both substrate consumption phases. Table I presents the performance figures for the best PLS model found for biomass, glycerol, acetate and plasmid concentrations. High accurate PLS regression models were achieved for biomass and plasmid, with a R^2 of 0.99 and a low RMSEP of 0.40 and 0.81, respectively. The biomass model produced a similar R^2 compared to previous reports on *E. coli* cultures, but lower prediction errors (Table II). Accurate PLS regression models were also obtained for glycerol and acetate, yielding a $R^2 \geq 0.93$ and a RMSEP of 0.32 and 0.06, respectively. The acetate prediction in terms of R^2 is more accurate compared to previous results on cultures of other bacteria (Table II).

TABLE I. Best PLS regression models for biomass, plasmid, glycerol and acetate concentration in terms of R^2 and RMSEP determined for a batch culture conducted with 6g/l of glycerol. The pre-processing technique, and the number of latent associated variables (LV) required to reach the minimal RMSEP is also indicated.

	Pre-processing	LV	R^2	RMSEP
Biomass (g/l)	no-preprocessing	10	0.99	0.40
Plasmid (mg/l)	constant off-set elimination	10	0.99	0.81
Glycerol (g/l)*	constant offset elimination	3	0.93	0.32
Acetate (g/l)*	constant offset elimination	3	0.95	0.06

* LOO validation.

All the best PLS regression models built did not use the second derivative as pre-processing, as opposed to other authors using *in-situ* NIR probes. For example, for the biomass model no pre-processing technique was necessary, and a unique constant offset elimination was applied for glycerol, acetate and plasmid models. This could be in part explained by the use of a Fourier Transform spectrometer, presenting an increased wavelength accuracy, spectral quality

and reproducibility in relation to dispersive systems, and by the use of a NIR probe working in transflectance mode, and presenting a mirror with a conical shape that avoids the accumulation of solids and air bubbles in the pathlength.

TABLE II. PREDICTIVE PERFORMANCE OF PLS MODELS OBTAINED BY OTHER AUTHORS IN *ON-LINE* AND *IN-SITU* NIR ANALYSES USING FIBER OPTIC TRANSMISSION/TRANSFLECTANCE PROBES.

Reference	Microorganism/ product	Predictive PLS performance for Biomass	Predictive PLS performance for Acetate
Arnold et al. (2002)	*E. coli*/ Somatropin	R^2=0.95 RMSEP=1.39	-
Cimander & Mandenius (2002)	*E. coli*/ Tryptophan	R^2=0.99 RMSEP=0.56	-
Tamburini et al. (2003)	*Staphylococcus*	R^2=0.95 SEP=1.21	R^2=0.92 SEP=0.56
	Lactobacillus	R^2=0.95 SEP=1.15	R^2=0.88 SEP=0.57
Tosi et al. (2003)	*Staphylococcus xylosus*	R^2=0.95 SEP=0.80	R^2=0.90 SEP=0.61
	Lactobacillus fermentum	R^2=0.93 SEP=0.95	R^2=0.81 SEP=0.51
	Streptococcus Thermophylus	R^2=0.95 SEP=0.70	-
Navrátil et al. (2005)	*Vibrio cholera*/ cholera toxin	R^2=0.99 SEP=0.2	R^2=0.92 SEP=0.28

For optimization purposes it is relevant that the PLS models are valid for a wide range of cultivation conditions. For that reason, three different cultivations based on different batch phases were combined for model building, conducted with three different mixtures of glucose and glycerol as carbon sources. For two of the three cultivations (B and C), the end of the batch phase (which coincides with the acetate consumption) was followed by an exponential feeding phase with glucose (Fig. 1).

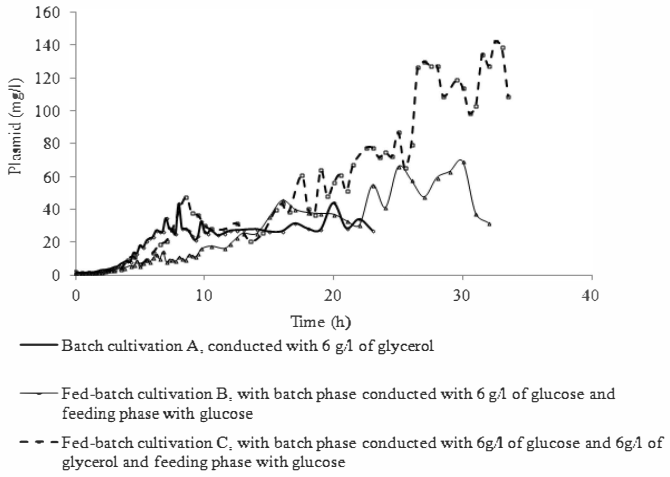

Figure 1. Plasmid production along three different cultures, conducted during the batch phase with different mixtures of glucose and glycerol, and during the feeding phase with glucose.

The three cultures studied, representing three different batch phases and two feeding phases, presented highly different plasmid productions per biomass. During the *E. coli* growth phases, batch culture A, conducted with glycerol, presented a specific plasmid production per biomass of 4.7 mg/g; fed-batch cultivation B, conducted with glucose on the batch phase, presented a specific plasmid production of 2.3 mg/g; and fed-batch cultivation C, conducted with a mixture of glucose and glycerol in the batch phase, presented a specific plasmid production of 6.3 mg/l. Furthermore, during the growth stationary phase, there were always high plasmid productions on all three cultures (data not shown). In spite of the distinct plasmid productions per biomass observed, a good predictive PLS model was build for the plasmid key variable, yielding an R^2 of 0.97 and a RMSEP of 7.5 (Fig. 2, Table III).

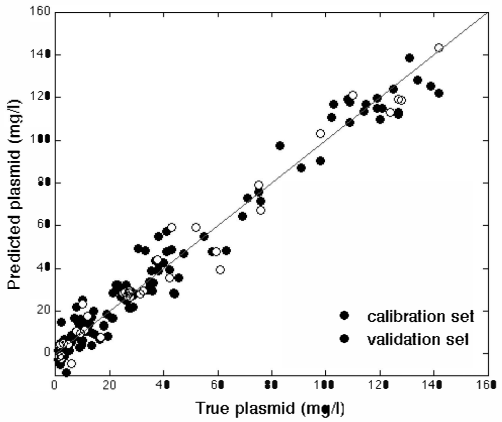

Figure 2. Experimental and predicted plasmid concentrations obtained by the PLS regression model based on the NIR spectra, considering data from three batches and two feeding modes of cultivation.

TABLE III. BEST PLS REGRESSION MODELS FOR BIOMASS, PLASMID, GLUCOSE, GLYCEROL AND ACETATE CONCENTRATION IN TERMS OF R^2 AND RMSEP, CONSIDERING DATA FROM THREE BATCHES AND TWO FEEDING MODES OF CULTIVATION. THE PRE-PROCESSING TECHNIQUE, AND THE NUMBER OF LATENT ASSOCIATED VARIABLES (LV) REQUIRED TO REACH THE MINIMAL RMSEP IS ALSO INDICATED.

	R^2	LV	RMSEP	Pre-processing
Biomass (g/l)	0.99	10	0.34	no pre-processing
Plasmid (mg/l)	0.97	10	7.52	SNV
Glucose (g/l)*	0.93	7	0.46	SNV
Glycerol (g/l)*	0.93	5	0.33	SNV
Acetate (g/l)	0.96	10	0.32	min-max normalization

* LOO validation.

The PLS model obtained for plasmid using different cultivations produced, as expected, slightly lower R^2 and higher RMSEP in relation to the use of a unique defined culture condition (Tables I and III). However, the RMSEP obtained is still from the same order of magnitude in relation to the conventional assay, based on plasmid extraction by alkaline cell lyses and HPLC analysis, i.e., 6%. Regarding the key variables biomass and acetate, the performance of the prediction models developed when using so different cultivation conditions increased in relation to the use of a defined cultivation condition, most probably due to the use of

a wider range of concentrations. In culture A, used for building PLS models presented in Table I, the maximum biomass and acetate concentrations were 5.6 and 3.3 g/l, respectively. Considering the three batches and the two feeding modes combined, the biomass and acetate range reached 15.5 and 5.3 g/l, respectively. The performance of the best PLS model obtained for the carbon source glycerol remains accurate, yielding similar R^2 and RMSEP values, even when using so distinct cultivation conditions and strategies. The PLS model developed for glucose over so different media compositions presented a high R^2 of 0.93 and a low RMSEP of 0.46 (Table III).

This work provides the tools for the development of *in-situ* monitoring techniques based on NIR spectroscopy for the plasmid bioproduction process, to reach an accurate real-time characterization of the key variables of the process, over batch and fed-batch cultivation strategies, and using different mixtures of the carbon sources glucose and glycerol. As the PLS models developed accurately predict for a wide range of culture conditions, the time spent in *off-line* analyses can therefore be avoided, enabling speeding the optimization procedure, while promoting process reproducibility and consistency in the final therapeutic product.

ACKNOWLEDGEMENT

This work was supported by the PTDC/BIO/69242/2006 project from the Portuguese Foundation for Science and Technology (FCT). M. B. Lopes gratefully acknowledges financial support from FCT (SFRH/BPD/73758/2010).

REFERENCES

[1] D. Bower and K. Prather, "Engineering of bacterial strains and vectors for the production of plasmid DNA", Appl. Microbiol. Biotechnol., vol. 82, pp. 805-813, 2009.

[2] Food and Drug Administration, "Guidance for Industry: PAT – A framework for innovatine pharmaceutical development, manufacturing and quality assurance", Washington DC, USA, 2004.

[3] European Medicines Agency, "Biologics working party (BWP), Process Analytical Technology (PAT) group and Industry", London, UK, 2007.

[4] T. Silva, P. lima, M. Roxo-Rosa, S. Hageman, L. P. Fonseca, and C. R. C. Calado, "Prediction of dynamic plasmid production by recombinant *Escherichia coli* fed-batch cultivations with a generalized regression neural network", Chem. Biochem. Eng. Q., vol. 23, pp. 419-427, 2009.

[5] T. Scholz, V. V. Lopes, and C. R. C. Calado, "High-throughput analysis of the plasmid bioproduction process in *Escherichia coli* FTIR spectroscopy", Biotechnol. Bioeng, vol. 109, pp. 2279-2285, 2012.

[6] M. Navrátil, A. Norberg, L. Lembrén, and C.-F. Mandenius, "On-line multi-analyzer monitoring of biomass, glucose and acetate for growth rate control of *Vibrio cholera* fed-batch cultivation", J. Biotechnol., vol. 115, pp. 67-79, 2005.

[7] S. Tosi, M. Rossi, E. Tamburini, G. Vaccari, A. Amaretti, and D. Matteuzzi, "Assessment of in-line near-infrared spectroscopy for continuous monitoring of fermentation processes", Biotechnol. Prog., vol. 19, pp. 1816-1821, 2003.

[8] E. Tamburini, G. Vaccari, S. Tosi, and A. Trilli, "Near-infrared spectroscopy: a tool for monitoring submerged fermentation process using an immersion optical-fiber probe", Appl. Spectrosc., vol. 57, pp. 132-138, 2003.

[9] S.A. Arnold, R. Gaensakoo, L.M. Harvey, and B. McNeil, "Use of at-line and in-situ near-infrared spectroscopy to monitor biomass in an industrial fed-batch *Escherichia coli* process", Biotechnol. Bioeng., vol. 80, pp.405-413, 2002.

[10] C. Cimander and C.-F. Mandenius, "Online monitoring of a bioprocess based on a multi-analyser system and multivariate statistical process modeling", J. Chem. Technol. Biotechnol., vol. 77, pp. 1157-1157, 2002.

[11] Y. Cai, S. Rodriguez, R. Rameswaran, R. Draghia-Akli, R. J. Juba, and H. Hebel, "Production of pharmaceutical-grade plasmids at high concentration and high supercoiled percentage", Vaccine, vol. 28, pp. 2046-2052, 2010.

[12] M. K. Danquah and G. M. Forde, "Growth medium selection and its economic impact on plasmid DNA production", J. Biosci. Bioeng., vol. 104, pp. 490-497, 2010.