Modelling, Monitoring and Control of Plasmid Bioproduction in *Escherichia coli* Cultures

Marta B. Lopes[1,2,3], Teresa Scholtz[1,2], Daniel Silva[1,4], Inês Santos[1], Tito Silva[1], Pedro Sampaio[1], Andreia Couto[1], Vitor V. Lopes[2,*], Cecília R.C. Calado[1,ξ]

[1]Engineering Faculty, Catholic University of Portugal, Rio de Mouro, Portugal
[2]Energy Systems Modelling and Optimization Unit, National Laboratory for Energy and Geology, Lisbon, Portugal
[3]Instituto de Telecomunicações, Instituto Superior Técnico, Lisboa, Portugal
[4]Instituto de Medicina Molecular, Lisboa, Portugal
[*] vitor.lopes@lneg.pt / [ξ] c.calado@fe.lisboa.ucp.pt (P.I.)

Abstract — **An integrated approach for modelling, monitoring and control the plasmid bioproduction in *Escherichia coli* cultures is presented. In a first stage, by the implementation of a kinetic model for *E. coli* cultures, a better bioprocess understanding was reached, concerning the availability of nutrients and products along the bioprocess, and their effects on the plasmid production. Results presented may provide significant help for future modelling and monitoring implementation. In a second stage, FTIR spectroscopy coupled with chemometrics, namely PLS regression, shows its potential as a high-throughput technique for simultaneously estimating the key variables involved in the plasmid production process by *E. coli* cultures run under distinct conditions. Finally, owing to on-line monitoring and process control, an NIR fibre optic probe and chemometrics provided promising results concerning the control of biomass and carbon sources in *E. coli* cultures.**

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

I. INTRODUCTION

In the last decade plasmid vectors have become extremely attractive for use in DNA vaccines and gene therapy, as they offer multiple advantages over viral vectors, namely, large packaging capacity, stability without integration and reduced toxicity [1]. On a small scale, plasmids are viewed as relatively easy to produce and purify. However, the industrial plasmid production with recombinant *Escherichia coli* cells is a complex process affected by many parameters, such as, medium composition, culture conditions and culture strategy, all playing an important role in the plasmid stability, plasmid copy number, and the amount of biomass produced [2].

An example of the complex interrelationship present in the plasmid expression system is the high impact that slight different compositions in the carbon source composed of mixtures of glucose and glycerol may present in the final plasmid copy number per cell, and consequently, on the culture productivity. Glucose is the main carbon source used to promote the bacterial growth. However, a primary barrier to total productivity in recombinant *E. coli* cultivation is the glucose metabolism to acetic acid, which can inhibit growth and decrease the recombinant product yield [3]. Two acetate production mechanisms under aerobic growth may occur in *E. coli*. The first, when the maximum oxygen transfer capacity of the reactor is reached, causing anaerobiosis; in the second, acetate is formed aerobically when the uptake of the carbon substrate is greater than its conversion to biomass and CO_2. Both mechanisms could be very important in aerobic cultivations in large-scale due to substrate and oxygen gradients [4]. The theoretical main advantage of using glycerol instead of glucose is that the former carbon source usually gives rise to low levels of acetic acid. However, in plasmid bioproduction systems, the production of high acetate concentrations using glycerol has been registered, and associated to low specific growth rates and high plasmid productions per biomass [1].

In order to analyze the impact of the above relationships on the highly complex *E. coli* cellular network it is relevant to develop models that will reproduce these relationships and allow the identification of the key variables of the process, and consequently help optimizing the medium and process conditions. Also of paramount importance, the development of monitoring techniques will provide information concerning the previously defined key variables in a high-throughput mode and/or in a real time frame.

Accessing real-time information on all variables involved in industrial bioprocesses is highly desired. On-line bioprocess measurements are currently mainly restricted to dissolved oxygen, pH and temperature. The concentrations of biomass, nutrients and products are usually obtained by off-line methods, consisting of removing the samples from the bioreactor and analyzing it by labor intensive techniques, *e.g.,* high performance liquid chromatography (HPLC).

Fourier transform infrared (FTIR) spectroscopy is a rapid off-line (*i.e.*, at-line) method that can be used as an alternative to the laborious off-line methods. Since each FTIR spectrum contains a "molecular fingerprint" of the sample being analyzed, it contains both quantitative and qualitative information about the sample composition. Multivariate calibration methods, such, as partial least squares (PLS) regression, have been successfully used for the prediction of the main compounds involved in many bioprocesses [5-7].

Modern FTIR equipments allow the simultaneous spectral acquisition from several samples provided by distinct bioprocesses, which makes FTIR spectroscopy highly promising for high-throughput analysis in bioprocess monitoring.

Near-infrared (NIR) spectroscopy is a rapid and non-invasive technique that enables the simultaneous measurement of several analytes along the bioprocess in real time. Despite being less informative that FTIR, providing broader and the fundamental vibration modes of molecules, the NIR weaker and overlapping bands can be handled by multivariate calibration techniques, which makes NIR far more attractive than FTIR, given its real-time, non-destructive operating way.

Many studies have already shown the potential of NIR for at-line analysis [6,8-12]. With respect to the on-line NIR monitoring, it can be performed either "ex-situ" or "in-situ". In "ex-situ" monitoring an NIR fibre optic probe can be mounted in a flow-through cell connected to the bioreactor [5,13] or placed in bioreactor windows [14,15]; for "in-situ" on-line monitoring the fibre optic probe is placed inside the bioreactor vessel. In the first case, however, samples in flow-through cells may not preserve all characteristics of the medium culture in the bioreactor (*e.g.*, agitation and aeration), thus not being fully representative of the bioprocess. The possibility of on-line monitoring an industrial fed-batch *E. coli* process using an NIR fibre optic probe ("in-situ") has already been demonstrated by Arnold et al. (2002)[16]. In that study good predictive models were obtained for biomass.

II. Goals

The present work studies the plasmid expression system *E. coli* DH5α containing the plasmid model pVAX-LacZ (Invitrogen) and focuses on the following three main goals:

- Development of an interpretative model for the bioproduction of plasmid, substrates and products in *E. coli* cultures. The model should be applicable in a wide range of media compositions and culture strategies, with special emphasis on those producing a high number of plasmid copies per cell, of great interest for control and optimization purposes;
- Development of high-throughput monitoring techniques based on FTIR spectroscopy (off-line) to enable the rapid and accurate characterization of the major variables provides by multi-cultures, by using for example multi-micro-bioreactors platforms; the plasmid expression profile will be therefore rapid characterized for a wide range of media compositions and conditions;
- Development of on-line monitoring techniques based on NIR spectroscopy to reach a better process understanding in real time, and consequently, robustness in control and optimization.

III. Development of a Kinetic Model for the Plasmid Bioproduction

The main goal of this task is the development of an interpretative kinetic model for batch cultivations of *E. coli* DH5α producing the plasmid model pVAX-LacZ, conducted in economic complex non-selective media containing mixtures of glucose and glycerol presenting high plasmid yields and productivities. The kinetic model describing the growth of *E. coli* is proposed based on the theoretical background and experimental results. The mass balance equations were formulated as follows:

$$\frac{dS}{dt} = -r_S, \frac{dG}{dt} = -r_G, \frac{dA}{dt} = r_{PA} - r_A, \frac{dX}{dt} = r_X, \frac{dP}{dt} = r_P,$$

where r_S, r_G and r_A are the rate of consumption of the substrates glucose, glycerol and acetate, and r_{PA}, r_X and r_P are the rate of production of acetate, biomass and plasmid, respectively. The above rates are a function of the specific growth rates (μ) for each substrate, as shown in Table I. The specific growth rate, μ, was based on the Monod equation, given by

$$\mu = \frac{\mu_{max} S}{K_S + S}, \quad (1)$$

where S is the concentration of the limiting substrate, μ_{max} is the maximum specific growth rate and K_S is a half saturation coefficient. Given the current knowledge about the biological kinetics, the above equation has been modified to account for inhibition factors, namely, the inhibition of the consumption of one substrate by the presence of others (Ki_{S-A}, Ki_{G-S}, Ki_{G-A}, Ki_{A-S}, Ki_{A-G}). The inhibitions of biomass by the acetate (Ki_{X-A}) and plasmid productions (Ki_{X-P}), as well as the inhibitions of plasmid by the acetate (Ki_{P-A}) and the specific growth rate ($Ki_{P-\mu}$), were also considered. A summarization of all model equations can be found in Table I.

TABLE I. Model Kinetic Equations

μ	$\mu_S + \mu_G + \mu_A$
μ_S	$\frac{\boldsymbol{\mu max_S}\, S}{\boldsymbol{K_S} + S} \frac{\boldsymbol{Ki_{S-A}}}{\boldsymbol{Ki_{S-A}} + A} \frac{\boldsymbol{Ki_{X-P}}}{\boldsymbol{Ki_{X-P}} + \frac{P}{X}} \frac{\boldsymbol{Ki_{X-A}}}{\boldsymbol{Ki_{X-A}} + A}$
μ_G	$\frac{\boldsymbol{\mu max_G}\, G}{\boldsymbol{K_G} + G} \frac{\boldsymbol{Ki_{G-S}}}{\boldsymbol{Ki_{G-S}} + S} \frac{\boldsymbol{Ki_{G-A}}}{\boldsymbol{Ki_{G-A}} + A} \frac{\boldsymbol{Ki_{X-P}}}{\boldsymbol{Ki_{X-P}} + \frac{P}{X}} \frac{\boldsymbol{Ki_{X-A}}}{\boldsymbol{Ki_{X-A}} + A}$
μ_A	$\frac{\boldsymbol{\mu max_A}\, A}{\boldsymbol{K_A} + A} \frac{\boldsymbol{Ki_{A-S}}}{\boldsymbol{Ki_{A-S}} + S} \frac{\boldsymbol{Ki_{A-G}}}{\boldsymbol{Ki_{A-G}} + G} \frac{\boldsymbol{Ki_{X-P}}}{\boldsymbol{Ki_{X-P}} + \frac{P}{X}} \frac{\boldsymbol{Ki_{X-A}}}{\boldsymbol{Ki_{X-A}} + A}$
r_S	$\frac{\mu_S}{\boldsymbol{Y_{X/S}}} X$
r_G	$\frac{\mu_G}{\boldsymbol{Y_{X/G}}} X$
r_A	$\frac{\mu_A}{\boldsymbol{Y_{X/A}}} X$
r_{PA}	$\left(\boldsymbol{Y_{PA/S}}\, \mu_S + \boldsymbol{Y_{PA/G}}\, \mu_G\right) X$
r_P	$\left(\boldsymbol{Y_{P/S}}\, \mu_S + \boldsymbol{Y_{P/G}}\, \mu_G + \boldsymbol{Y_{P/A}}\, \mu_A\right) \frac{\boldsymbol{Ki_{P-\mu}}}{\boldsymbol{Ki_{P-\mu}} + \mu} \frac{\boldsymbol{Ki_{P-A}}}{\boldsymbol{Ki_{P-A}} + A} X$

(The parameters estimated by the model are presented in bold).

Three *E. coli* cultures, run under mixtures of glucose and glycerol as carbon sources, and presenting distinct biomass growth and plasmid copy number per cell, were used in this study (Fig. 1). Details on the pre-culture and cultivation procedures, as well as the reference analyses performed for the determination of biomass, glucose, glycerol, acetate and plasmid concentrations, can be found in Martins (2008)[17].

The proposed model was applied to the single cultures and to the three cultures simultaneously. The parameters of the model were estimated by non-linear least squares numerical optimization, using the Levenberg-Marquardt algorithm [18]. Model predictions for the concentrations of each variable when applying the model to each culture separately, are depicted in Fig. 1. Overall, very good estimates were obtained, with very small errors of prediction. The prediction error values substantially increased when trying to model the three cultures together. The distinct initial conditions, generating distinct biochemical processes, make difficult a unique model to describe all glucose-glycerol *E. coli* growth environments.

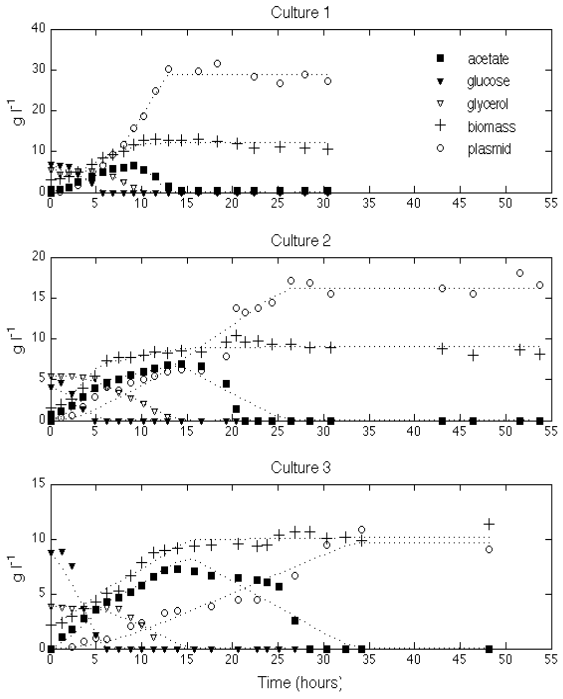

Figure 1: Experimental concentrations (symbols) and estimated concentrations (dotted lines) obtained by the full model applied to each single culture. To each culture, different parameters constants were observed. The plasmid concentration (mg l^{-1}) is divided by a constant c = 3.

The estimated value for the inhibition constants formulated by the model revealed the absence of Ki_{S-A}, Ki_{G-A}, Ki_{A-S}, $Ki_{P-\mu}$, Ki_{P-A}, and Ki_{X-P} (with their contribution in the model kinetic equations approaching the unity). These inhibitions were excluded from the model. Both Ki_{A-G} and Ki_{G-S} are evident in Fig. 1, with the consumption of glycerol only starting when there is no more glucose in the medium, and the reduction of the concentration of acetate occurring when glycerol had been totally consumed. The model points out that the inhibition of the consumption of acetate by glucose (Ki_{A-S}) is irrelevant, indicating that cell is consuming acetate while consuming glucose, but at a lower velocity compared to the velocity of acetate production, and therefore an acetate accumulation net is observed in medium along the glucose consumption phase. An equivalent consumption of acetate while glycerol is being consumed by the cell can also be inferred, given the not so strong inhibition of the consumption of acetate by the presence of glycerol (Ki_{A-G}). The reduced model was based on the full-model but without the previous pointed inhibition factors. The reduced model presented to the resulting seventeen parameters the same behavior as the full model, when considering the single cultures apart or the three distinct cultures all together to estimate the parameters, without any cost for the objective function. The reduced model was thus taken for further analyses.

Given the inherent difficulty of simultaneously modelling distinct experimental conditions, a novel modelling strategy was developed based on the identification of a set of parameters that are common to all experimental conditions, designated by common parameters, and a set of parameters, whose values are experiment-dependent, designated by free parameters. This way, the common parameters could be simultaneously estimated and the free parameters estimated in a case by case fashion. The free parameters might explain the differences between cultures and provide crucial knowledge about the cell behavior, thus fostering guidelines for model improvement and process scale-up. An iterative search was performed to identify and fix the estimated parameters that were common to the three total distinct cultures, in a forward selection mode based on stepwise regression. Starting from the reduced model estimating all seventeen parameters in the three cultures, with all parameters as free (culture-dependent), the first parameters to be tested as common were those showing very similar values among cultures. *F*-tests using a 0.05 significance level were performed to compare the reduced model estimating p parameters and a restricted model estimating $p-(N-1)$ parameters (N is the number of experiments). Each parameter was defined as common when the restricted model provided a statistically better fit compared to the full model. The procedure ended when all parameters had been tested.

Eleven parameters were fixed using the common parameter strategy, leaving six parameters that are culture-dependent (the maximum specific growth μmax_S, μmax_G and μmax_A and the biomass and plasmid yields $Y_{X/S}$, $Y_{X/G}$ and $Y_{P/A}$). Predictions of the concentrations of each metabolite using the above strategy produced very similar plots to the ones obtained using all free parameters (figure not shown). This result shows that only six free, culture-dependent parameters, are in the basis of the errors obtained by the model.

A crucial role of the acetate on the plasmid production was found. Distinct estimates of the plasmid yield on acetate ($Y_{P/A}$) were obtained for the three cultures (Table II). A slight increase in $Y_{P/A}$ occured during the glucose and glycerol uptake stages, then achieving its maximum in the descending phase of the acetate consumption phase, when the growth rate is constant. The high impact of acetate in the plasmid production rate was also observed, being almost entirely depended from the acetate, followed by a very small contribution from glycerol, and an almost absent contribution from glucose (results not shown).

This study provided important knowledge on the biological kinetics of the *E. coli* growth in glucose-glycerol environments. A model describing the *E. coli* kinetics enabled the possibility of fixing eleven parameters and leaving six that are culture-dependent. This is in fact a tremendous progress towards the implementation of a model describing a high

concentration range of glucose-glycerol growth environments. The present results may significantly contribute to the optimization and scale-up of batch glucose-glycerol cultivations for the plasmid production.

TABLE II. ESTIMATES OF THE CULTURE-DEPENDENT PARAMETERS OBTAINED FOR THE THREE CULTURES

Parameter	μmax_S	μmax_G	μmax_A	$Y_{X/S}$	$Y_{X/G}$	$Y_{P/A}$
Culture 1	0.193	0.082	0.003	0.895	0.664	124.991
Culture 2	0.313	0.050	0.001	1.070	0.675	78.473
Culture 3	0.188	0.058	0.001	0.421	1.238	39.962

IV. HIGH-THROUGHPUT PLASMID BIOPROCESS MONITORING BASED ON FTIR-SPECTROSCOPY

The present task presents a methodology based on FTIR spectroscopy and the use of chemometrics for a rapid automatable high-throughput analysis of the plasmid bioproduction process in *E. coli*. For this study, five batch cultures were used (pre-culture and cultivation procedures described elsewhere [17]), run under different initial medium compositions. The goal was to reach different biomass and plasmid production behaviors, with the maximum plasmid and biomass concentrations varying from 11 to 95 mg l^{-1} and 6.8 to 12.8 g l^{-1}, respectively, and resulting in plasmid productions per biomass between 0.4 and 5.1 mg g^{-1}.

Samples from the culture broth were taken from the bioreactor along the culture growth, and immediately centrifuged, to reduce spectral interference from the nutrients, such as complex N-sources, present in the culture broth. 30 µl of resuspended pellet in NaCl 0,9% (w/v) were placed on IR-transparent Zn-Se microliter plates with 96 wells (Bruker Optics, Germany) and subsequently dehydrated for 2.5 hours in a vacuum desiccator (ME2, Vaccubrand, Germany). The FTIR spectra were obtained by averaging 40 scans from 4000 to 400 cm^{-1} by a *HTS-XT* associated to *Vertex-70* spectrometer (Bruker Optics, Germany) at a resolution of 2 cm^{-1}.

PLS regression was used to relate the spectra to the known reference values for plasmid and biomass, obtained by reference methods described in Silva et al. (2009)[1]. The number of latent variables (LV) to be incorporated into the PLS models was chosen based on the mean squared error of prediction (MSEP) estimated by a leave-one-out (LOO) cross-validation. The predictive performance of the PLS models was evaluated by the coefficient of determination (r^2) between the experimental and predicted concentration and the MSEP. Both values were calculated based on a LOO cross-validation. First derivatives using a Savitsky-Golay filter with a 15-point window and a second order polynomial fit were used as pre-processing technique, in order to remove baseline drifts from spectra. Moreover, a Monte-Carlo strategy to identify the spectral regions which are relevant for the plasmid and biomass models was employed, as it is known that modelling can be improved by excluding spectral areas which do not contain analyte specific information [5]. For this purpose, the spectra were divided into nineteen equally sized intervals. In order to determine which of these intervals were related to the biomass and plasmid concentrations, 35000 PLS models randomly including some of these intervals were established and the MSEP was assessed by LOO cross-validation. The MSEP of the best 50 and worst models, along with the wavenumber selection employed, were saved.

The large impact of wavenumber selection on the prediction of the plasmid concentration can be seen in the resulting MSEPs varying from 39 to 236 mg l^{-1}. For the plasmid model, the spectral regions between 590-1130 cm^{-1}, 1670-2025 cm^{-1} and 2565-3280 cm^{-1}, were found to be highly relevant, as they were incorporated in over 75% of the best models. For the biomass, the best wavenumber selections were between 900-1200 cm^{-1}, 1500-1800 cm^{-1} and 2850-3200 cm^{-1}, also incorporated in 75% of the best models. By using the wavenumber regions of the first derivative spectra selected by the Monte-Carlo approach, coefficients of determination of 0.91 for the plasmid (MSEP=39 mg l^{-1}) and 0.89 (MSEP=1.1 g l^{-1}) for the biomass concentration, could be obtained (Table III, Fig. 2), which corresponds to an improvement in the PLS models' performance of about 20 and 30%, respectively.

The collected spectra contain direct information about the dry cell weight and plasmid concentration, since they were acquired from the biomass of the cultures. However, due to the cell metabolism, there can also be correlations between the absorbance measurements and the concentration of the nutrients and metabolites in the medium. In this work, these correlations were used to predict the concentration of glucose, glycerol and acetic acid in the culture broth. For this purpose, PLS regression models were built on the whole first derivative spectra, the results are shown in Table III.

TABLE III. PREDICTION PERFORMANCE OF THE PLS MODELS FOR THE CONCENTRATION OF GLUCOSE, GLYCEROL AND ACETIC ACID, OBTAINED BY HIGH-THROUGHPUT FTIR SPECTROSCOPY ANALYSIS.

	Number of samples	LV[c]	r^2	MSEP
Glucose	36	5	0.85	1.1
Glycerol	84	15	0.86	0.6
Acetic acid	114	11	0.76	0.8
Acetic acid-C[a]	69	6	0.92	0.3
Acetic acid-P[b]	45	10	0.85	0.5
Biomass	114	12	0.89	1.1
Plasmid	114	13	0.91	39

[a]Acetic acid in ascending phase.
[b]Acetic acid in descending phase.
[c]Number of latent variables used.

For the prediction of acetic acid, a smaller coefficient of determination was found (r^2 = 0.76). Acetic acid is produced by the bacteria during the glucose and glycerol consumption phases, and mainly consumed when neither glucose nor glycerol is available, which leads to changes in the bacteria's metabolism. Moreover, all cultures were conducted with the same composition of nitrogen source, and consequently different ratios of C/N. Thus, at the end of the cultures, when high biomass concentrations are obtained, severe nutrient limitations concerning micronutrients present in the nitrogen source are expected, with these nutrient limitations being highly different for the distinct cultures. Therefore, the bacteria cell composition may drastically differ among the

cultures, which can explain the poorer predictive capacity of the acetate model. To confirm this, two models were built for the acetate, one for the acetate ascending phase, and another for the acetate descending phase. Higher coefficients of determination and lower prediction errors were obtained.

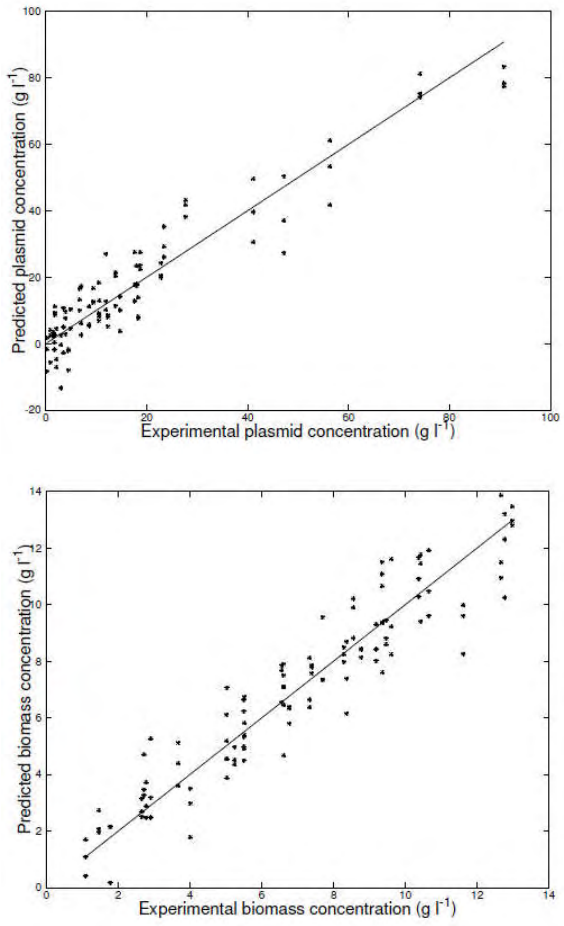

Figure 2: Measured and predicted concentration values for plasmid and biomass obtained by PLS regression, obtained by high-throughput FTIR spectroscopy analysis.

This task shows the potential of FTIR spectroscopy coupled to PLS regression for high-throughput analysis in bioprocess monitoring. The concentrations of plasmid and biomass, which are productivity determining factors, could accurately be estimated during *E. coli* bioprocesses. Moreover, even though the spectra were collected from the culture biomass and not the whole culture broth, it was possible to determine the concentration of the carbon sources glucose and glycerol and the by-product acetic acid in the culture medium. The good prediction models obtained from the FTIR spectroscopy analysis, the short analysis time, and the possibility of simultaneously estimating in a high-throughput way plasmid, biomass, nutrients, products and by-products from multi-cultures conducted in distinct conditions, turns even more attractive the multi-bioreactor platforms as an extremely useful tool in the plasmid bioprocess development.

V. On-line Bioprocess Monitoring Based on NIR Spectroscopy

Preliminary studies were conducted to explore the NIR capacity for predicting the *E. coli* DH5α biomass concentration and the carbon source glycerol (initial concentration 6 g l^{-1}) in an *E. coli* batch process. Pre-culture and cultivation procedures are described in Martins (2008)[17]. Samples were taken from the bioreactor every 15 minutes during the glycerol uptake phase, and hourly hereafter, for biomass determination (using both optical density at 600 nm and dry cell weight) and glycerol quantification by HPLC. NIR measurements were obtained using an NIR fibre optic probe *IN-271P* (Bruker Optics, Germany) coupled to a *Vertex-70* spectrometer (Bruker Optics, Germany). NIR spectra (Fig. 3) were collected every 2 minutes in the 12500-5400 cm^{-1} range, with 8 cm^{-1} resolution. Each spectrum consisted of 32 coadded scans.

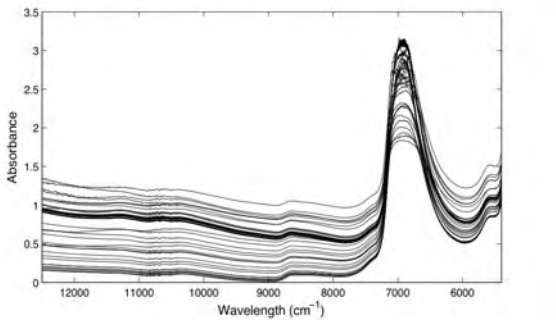

Figure 3: Near-infrared spectra of samples along the bioprocess obtained by on-line NIR spectroscopy analysis (spectral window 12500 and 5400 cm^{-1}).

Partial least squares models were built for biomass and glycerol. A number of pre-processing techniques were explored in order to increase the model's predictive performance, by removing physical phenomena from the spectra, namely, constant offset elimination, min-max normalization, standard normal variate (SNV), multiplicative scatter correction (MSC) and spectral derivatives. Wavelength selection was also performed to identify the spectral regions with relevant information on the variables under study. The procedure consisted of dividing the spectral window into 10 subregions and choosing the combination of regions and pre-processing providing the smallest prediction error.

TABLE IV. Summary of the Measures of Performance of the PLS Regression Models Built for Biomass and Glycerol

	r^2	LV[a]	RMSEP	NIR region (cm^{-1})	Pre-processing
Biomass (OD 600 nm)	0.99	10	0.4	12493.2-11073.8; 10368-9654.4; 8238.8-7525.3; 6819.4-6105.8	no pre-processing
Biomass (dry cell weight)	0.99	6	0.2	10368-8235	constant offset elimination
Glycerol	0.94	3	0.3	9656.3-7525.3	constant offset elimination

[a]Number of latent variables used.

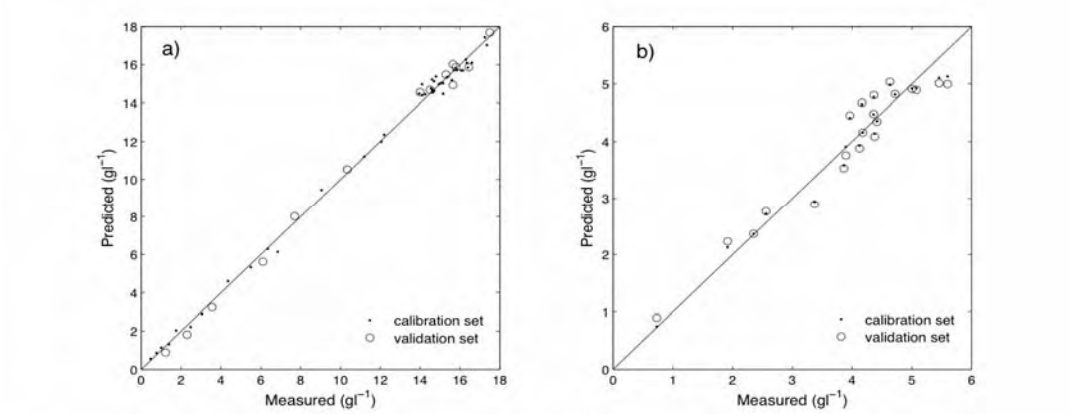

Figure 4: Measured and predicted concentration values for a) biomass (determined by dry cell weight) and b) glycerol obtained by PLS regression.

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 validation for the biomass, given the larger number of samples, and on LOO cross-validation for glycerol. A summary of each model measure of fit, along with the spectral window and pre-processing used for model building, can be found in Table IV. PLS modelling proved to be very accurate for both biomass measures, with an r^2 higher than 0.99 (Figure 4a, Table IV). Good prediction results were found for glycerol, with an r^2=0.94 (Figure 4b, Table IV).

Future research will focus on the prediction of the concentration of the plasmid pVAX-LacZ over the culture time course, as it has been shown to be successfully predicted by off-line FTIR analysis. The big advantage of using NIR instead of FTIR is the real-time measurement of the variables during the entire bioprocess, with no need for sample preparation.

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] 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.

[2] K.J. Prather, S. Sagar, J. Murphy, and M. Chartrain, "Industrial scale production of plasmid DNA for vaccine and gene therapy: plasmid design, production, and purification", Enzyme Microb. Technol., vol. 37, pp.865-883, 2002.

[3] W. Johnston, M. Stewart, P. Lee, and M. Cooney, "Tracking the acetate threshold using DO-transient control during medium and high cell density cultivation of recombinant *Escherichia coli* in complex media", Biotechnol. Bioeng., vol. 84, pp.314-323, 2003.

[4] B. Xu, M. Jahic, G. Blomsten, and S.-O. Enfors, "Glucose overflow metabolism and mixed-acid fermentation in aerobic large-scale fed-batch processes with *Escherichia coli*", Applied Microb. Biotech., vol. 51, pp.564-571, 1999.

[5] M. Kansiz, J.R. Gapes, D. McNaughton, B. Lendl, and K.C. Schuster, "Mid-infrared spectroscopy coupled to sequential injection analysis for the on-line monitoring of the acetone-butanol fermentation process", Anal. Chim. Acta, vol.438, pp.175-186, 2001.

[6] S. Sivakesava, J. Irudayaraj, and D. Ali, "Simultaneous determination of multiple comoponents in lactic acid fermentation using FT-MIR, NIR, and FT-Raman spectroscopic techniques", Process Biochem.., vol. 37, pp.371-378, 2001.

[7] P. Fayolle, D. Picque, and G. Corrieu, "Monitoring of fermentation processes producing lactic acid bacteria by mid-infrared spectroscopy", Vib. Spectrosc, vol. 14, pp.247-252, 1997.

[8] J.W. Hall, B. McNeil, M.J. Rollins, I. Drapper, B.G. Thompson, and G. Macaloney, "Near-infrared spectroscopic determination of acetate, ammonium, biomass, and glycerol in an industrial *Escherichia coli* fermentation", Applied Spec., vol. 50, pp. 102-108, 1996.

[9] S. Vaidyanathan, A. Arnold, L. Matheson, P. Mohan, G. Macaloney, B. Mcneil, and L. Harvey, "Critical evaluation of models developed for monitoring an industrial submerged bioprocess for antibiotic production using near-infrared spectroscopy", Biotechnol. Prog., vol. 16, pp. 1098-1105, 2000.

[10] S. Vaidyanathan, S.A. Arnold, L. Matheson, P. Mohan, B. McNeil, and L. Harvey, "Assessment of near-infrared spectral information for rapid monitoring of bioprocess quality", Biotechnol. Bioeng., vol. 74, pp.376-388, 2001.

[11] G. Macaloney, I. Draper, J. Preston, K.B. Anderson, M.J. Rollins, B.G. Thompson, J.W. Hall, and B. McNeil, "At-line control and fault analysis in an industrial high cell density *Escherichia coli* fermentation, using NIR spectroscopy", Trans IChemE, vol. 74, pp. C212-218, 1996.

[12] G. Macaloney, J.W. Hall, M.J. Rollins, I. Draper, K.B. Anderson, J. Preston, B.G. Thompson, and B. McNeil, "The utility and performance of near-infrared spectroscopy in simultaneous monitoring of multiple components in a high cell density recombinant *Escherichia coli* production process", Bioprocess Eng., vol. 17, pp. 157-167, 1997.

[13] G. Vaccari, E. Dosi, A.L. Campi, A. Gonzalez-Vara, D. Matteuzi, and G. Mantovani, "A near infrared spectroscopy technique for the control of fermentation processes: an application to lactic acid fermentation", Biotechnol. Bioeng., vol. 43, pp. 913-917, 1994.

[14] Z. Ge, A.G. Cavinato, and J.B. Callis, "Noninvasive spectroscopy for monitoring cell density in a fermentation process", Anal. Chem., vol. 66, pp. 1354-1362, 1994.

[15] A.G. Cavinato, D.M. Mayes, Z. Ge, and J. Callis, "Noninvasive method for monitoring ethanol in fermentation processes using fiber-optic near-infrared spectroscopy", Anal. Chem., vol. 62, pp. 1977-1982, 1990.

[16] 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.

[17] J.G. Martins, "Importância da fonte de carbono na produção de plasmídeos em culturas não selectivas e de elevada densidade celular de *Escherichia coli*", MSc Thesis in Biomedical Engineering, Engineering Faculty, Catholic University of Portugal, 2008.

[18] J.J. Moré, The Levenberg Marquardt algorithm: implementation and theory, In *Numerical Analysis*, G. A. Watson, editor, Lecture Notes in Mathematics , vol. 630, Springer, 1977, pp.105-116.