In Situ Near-Infrared (NIR) Versus High-Throughput Mid-Infrared (MIR) Spectroscopy to Monitor Biopharmaceutical Production

Kevin C. Sales,[a] Filipa Rosa,[a] Pedro N. Sampaio,[a] Luís P. Fonseca,[b] Marta B. Lopes,[a,c] Cecília R.C. Calado[d,*]

[a] *Engineering Faculty, Catholic University of Portugal, Estrada Octávio Pato, 2635-631, Rio de Mouro, Portugal*
[b] *IBB-Institute for Biotechnology and Bioengineering, Centre for Biological and Chemical Engineering, Instituto Superior Técnico, University of Lisbon, Av. Rovisco Pais, 1049-001 Lisboa, Portugal*
[c] *Institute of Telecommunications, Instituto Superior Técnico, Av. Rovisco Pais, 1049-001 Lisboa, Portugal*
[d] *ISE–Instituto Superior de Engenharia de Lisboa, Instituto Politécnico de Lisboa, Rua Conselheiro Emídio Navarro 1, 1959-007, Lisboa*

The development of biopharmaceutical manufacturing processes presents critical constraints, with the major constraint being that living cells synthesize these molecules, presenting inherent behavior variability due to their high sensitivity to small fluctuations in the cultivation environment. To speed up the development process and to control this critical manufacturing step, it is relevant to develop high-throughput and in situ monitoring techniques, respectively. Here, high-throughput mid-infrared (MIR) spectral analysis of dehydrated cell pellets and in situ near-infrared (NIR) spectral analysis of the whole culture broth were compared to monitor plasmid production in recombinant *Escherichia coli* cultures. Good partial least squares (PLS) regression models were built, either based on MIR or NIR spectral data, yielding high coefficients of determination (R^2) and low predictive errors (root mean square error, or RMSE) to estimate host cell growth, plasmid production, carbon source consumption (glucose and glycerol), and by-product acetate production and consumption. The predictive errors for biomass, plasmid, glucose, glycerol, and acetate based on MIR data were 0.7 g/L, 9 mg/L, 0.3 g/L, 0.4 g/L, and 0.4 g/L, respectively, whereas for NIR data the predictive errors obtained were 0.4 g/L, 8 mg/L, 0.3 g/L, 0.2 g/L, and 0.4 g/L, respectively. The models obtained are robust as they are valid for cultivations conducted with different media compositions and with different cultivation strategies (batch and fed-batch). Besides being conducted in situ with a sterilized fiber optic probe, NIR spectroscopy allows building PLS models for estimating plasmid, glucose, and acetate that are as accurate as those obtained from the high-throughput MIR setup, and better models for estimating biomass and glycerol, yielding a decrease in 57 and 50% of the RMSE, respectively, compared to the MIR setup. However, MIR spectroscopy could be a valid alternative in the case of optimization protocols, due to possible space constraints or high costs associated with the use of multi-fiber optic probes for multi-bioreactors. In this case, MIR could be conducted in a high-throughput manner, analyzing hundreds of culture samples in a rapid and automatic mode.

Index Headings: **Biopharmaceuticals; Bioprocess monitoring; Mid-infrared spectroscopy; Near-infrared spectroscopy; Partial least squares models; PLS: Process analytical technologies; PAT.**

INTRODUCTION

Innovative and effective biopharmaceutical medicines represent the cutting edge of the pharmaceutical sector, represent eight of the 20 top selling drugs and more than 40% of all pharmaceutical industry research and products in development in 2012, and account for more than US$165 billion worldwide in annual revenue.[1] Biopharmaceuticals are proteins or nucleic acids derived by the use of recombinant DNA, controlled gene expression methods, or a combination.[2] However, the process development for biopharmaceuticals presents several relevant constraints, with the major constraint being that these molecules are synthesized by living cells, presenting inherent variability concerning production capacity, further enhanced by sensitivity to the manufacturing environment.[2,3] In addition, biopharmaceuticals often exist in many different versions or variations of molecules (e.g., different glycosylation and folding patterns). Due to the large dimensions of the structures, which could reach up to 25 000 atoms, and their three-dimensional complexity, they are difficult to characterize based on biological functionality and security, e.g., concerning the immunogenicity profile.[2,3] Even a small uncontrolled change in the culture environment of the producing cell, for example medium composition, pH, temperature, and dissolved oxygen concentration (DOC), can alter cell metabolism, directly changing the biopharmaceutical activity and security profile.[2–5] Due to these characteristics, and with the aim to protect consumers, global regulatory agencies impose stringent quality guidelines that have been chang-

Received 9 May 2014; accepted 15 December 2014.
* Author to whom correspondence should be sent. E-mail: ccalado@deq.isel.pt.
DOI: 10.1366/14-07588

0003-7028/15/6906-0760/0

ing to meet the quality by design framework, a framework that encloses the scientific understanding of the whole life cycle of the products, including the manufacturing process.[6] The correct understanding of the bioprocess may be used to speed up the development process toward an optimized and more economical manufacturing process. Furthermore, it could enable identifying sources of variability and ultimately counteract the process variations by using advanced process controls. Indeed, the International Conference on Harmonisation (document Q8, Pharmaceutical Development)[7] suggests that all process parameters that influence the critical process attributes should be controlled. The same idea is present in the process analytical technologies (PAT) system introduced in 2004 by the U.S. Food and Drug Administration,[8] as it is used to design, analyze, and control the bioprocess, based on monitoring the critical quality attributes of raw materials and process along time with the final goal of ensuring a high-quality and regulatory-compliant product.[4,5,9,10]

A prerequisite for speeding up biopharmaceutical process development and bioprocess control is therefore the development of adequate monitoring techniques. However, there are only a few well-established sensors for on-line monitoring of parameters, e.g., temperature, pH, DOC, stirring speed, gas and liquid flow rates, and oxidation–reduction potential. The accurate monitoring of the remaining critical variables of the bioprocess (e.g., biomass, nutrients, inhibitory products, and product) is usually based on off-line analyses conducted on samples taken from the bioreactor and subsequently analyzed by multi-analytical techniques, such as enzymatic assays, high-performance liquid chromatography (HPLC), and immunoassays (e.g., enzyme-linked immunosorbent assay). These off-line analyses pose serious constraints, such as high cost, risk of contamination, and a delay between the sampling and the time of analysis. These characteristics are more relevant at industrial production scales than in optimization protocols, where the contamination of a large-scale bioreactor will have a serious economic impact, and the time delay to achieve critical information of the culture performance could impair the real-time control. At-line analysis speeds up the acquisition of critical information in relation to off-line analysis, in spite of still involving sampling; it implies a rapid analysis in proximity to the bioreactor.[10–14]

On-line analysis does not require culture sampling and can be conducted in situ, e.g., by using an analytical probe directly immersed in the culture broth, or ex situ. In ex situ monitoring, the analytical probe can be mounted in an external flow-through cell connected to the bioreactor or placed in bioreactor windows. These latter techniques are usually limited to glass vessels, as used in a laboratory scale. At an industrial scale, the steel bioreactors do not enable an adequate bioreactor window for this purpose, and a flow-through cell would critically increase the risk of contamination. Moreover, samples in the flow-through cells may not preserve all characteristics of the medium culture in the bioreactor, thus not being fully representative of the bioprocess. Therefore, the ideal monitoring technique for control purposes at large-scale productions should be conducted on-line and in situ; be automated; enable rapid acquisition of information (i.e., in real time), with minimal or no sample preparation and reagent consumption; and enable an accurate analysis of all critical variables of the bioprocess from a single measurement.[10–14] For optimization protocols (e.g., to screen the optimum combination of expression system, media composition, and cultivation strategy),[15–17] which are usually conducted with multi-bioreactors at a laboratory scale,[18–21] the at-line and high-throughput monitoring mode could present economic advantages over the in situ monitoring mode as a result of the costs associated with the in situ multi-probes. Fourier transform infrared spectroscopy (FT-IR), a highly sensitive technique that detects the vibrational modes of functional groups in molecules, is a very promising candidate to be applied either on-line and in situ, or at-line and in high-throughput mode.[11,12,22–26]

Depending on the infrared (IR) region used, MIR or near-infrared (NIR), IR spectroscopy presents specific characteristics and therefore specific advantages and limitations that ultimately may complement each other. MIR spectroscopy reflects the fundamental vibrations of the molecular bonds, whereas NIR spectroscopy reflects overtones and combinations of vibrations, thereby making MIR spectra more informative concerning the samples' biomolecular composition. Both NIR and MIR light can be transported using optical fibers. For example, NIR fiber optic probes have been mounted ex situ, either in an external flow-through cell connected to bioreactors[27] or in bioreactor windows,[28,29] or in situ to predict biomass, some nutrients, products such as proteins and plasmids, and by-products.[30–37] The use of ex situ analysis as previously pointed out is not advised for large-scale steel bioreactors. Some studies also detected interference in the in situ NIR analysis due to variations in temperature, airflow, and stirring speed at the bioreactor.[29,34–36] Indeed, during cultivation of microorganisms, a highly dynamic range of aeration (that could easily achieve 1.5 units of volume of air per volume of medium and per minute), agitation speed (that can easily range from 100 to more than 1000 rpm), and biomass concentration is normally achieved, as is the case in the present work. The majority of the above-mentioned limitations could be overcome by mathematical modeling or by using an NIR transflectance probe,[29,34–36] as presented by Lopes et al.[30] who monitored the plasmid bioproduction in recombinant *Escherichia coli* strains and by Sampaio et al.[31] who monitored a recombinant protein production in *Saccharomyces cerevisiae* strains. In both studies, a transflectance probe was used that combined transmittance and reflectance operation modes. In transflectance, the light passes through the sample twice, resulting in a blend of transmitted and reflected light, depending on the properties of the sample, which changes along the culture. Furthermore, some probes present a conical-shaped mirror that minimizes the accumulation of solids and air bubbles in the path length. In situ analysis using a MIR fiber optic probe has also been conducted with attenuated total reflection (ATR) detection mode,[23,24,38] based on light reflection through a crystal. An important disadvantage of this system, resulting from the MIR light characteristics, is that the fiber optic cable is generally

limited to a very few meters, whereas the NIR light can be led by fiber optic cables at larger distances, more applicable at large-scale productions.[11] An advantage of MIR spectroscopy for optimization protocols is the possibility of spectral acquisition in at-line and high-throughput transmission modes, using plates with multi-microwells that enable the simultaneous analysis of hundreds of micro-samples, only requiring the culture broth dehydration before analysis.[39,40]

This work aims to compare MIR and NIR spectroscopy for monitoring in high-throughput and in situ mode, respectively, a recombinant *E. coli* DH5 culture producing the plasmid pVAX-LacZ. The choice of the expression system model was influenced by *E. coli* being the most widely used microbial expression host and by the growing demand for pharmaceutical plasmid DNA products, while regulatory approval is being granted for non-viral human gene therapy and DNA vaccination.[41–45] The use of FT-IR to monitor plasmid bioproduction has been studied previously; for example, Scholz et al.[39] used high-throughput MIR spectroscopy and Lopes et al.[30] used in situ NIR spectroscopy to monitor plasmid bioproduction in *E. coli* cultures. However, the efficiency of these two techniques for monitoring plasmid bioprocesses cannot be compared, as different culture conditions were used. Moreover, few samples were used for model building in the MIR spectroscopy monitoring, potentially impairing the model prediction capability. Thus, in the present work, high-throughput MIR spectral analysis and in situ analysis using an NIR fiber optic probe were compared to monitor the same cultures. Several partial least squares (PLS) regression models based on MIR and NIR data from recombinant *E. coli* cultures conducted with different media conditions and different cultivation modes (batch and fed-batch) were built to estimate the critical variables of the bioprocess.

MATERIALS AND METHODS

Cultivation. *Escherichia coli* DH5-α containing the plasmid model pVAX-LacZ (Invitrogen, USA) was used. The stock cultures, grown on 2% (w/v) Luria broth (Sigma, UK) and 30 μg/mL kanamycin (Sigma-Aldrich, Germany), were maintained at −80 °C in 40% (v/v) glycerol solution (Panreac Quimica SA, Spain) with 10 mM Tris-HCl (Sigma-Aldrich) buffer (pH 8.0). Each inoculum culture was started with an aliquot of 10 μL of stock culture placed in a 1 L shake flask containing 300 mL with 20 g/L bactotryptone (BD, USA), 10 g/L yeast extract (Difco, USA), 10 g/L sodium chloride (NaCl) (Merck, Germany), and 30 μg/mL kanamycin. The inoculum culture was grown to mid-exponential phase (corresponding to an optical density at 600 nm of approximately 5) and then used to inoculate at 10% (v/v) the main batch culture.

The main batch cultivation was performed in a 2 L bioreactor (Biostat MD, B. Braun, Germany) with a 1.8 L working volume, in the absence of antibiotic. Cultivation was maintained at 7.0 ± 0.1 pH by automatic control through 1 M sodium hydroxide (Fluka, Switzerland) or 1 M hydrochloric acid (Sigma-Aldrich) addition and at 37 ± 0.1 °C with a minimal DOC of 30 ± 5% air saturation, by automatic adjustment of the agitation rate and adjustment of the airflow rate range between 1.0 and 1.5 volume of air/volume of medium/minute (vvm). The initial batch cultivation media of the three cultures studied contained 10 g/L yeast extract, 20 g/L Bacto-tryptone, and 7 g/L glycerol (culture A); 7 g/L glucose (culture B); and 6 g/L glycerol and 8 g/L glucose (culture C). An exponential feeding phase was started on cultures B and C with a feeding of 0.3 L of medium containing 22.5 g of yeast extract, 22.5 g of Bacto-tryptone, and 45 g of glucose, considering a maximum specific growth rate of 0.18 h^{-1} and a constant yield of biomass per glucose of 0.6 g/g as described previously.[46] Samples were taken from the bioreactor along the culture and subsequently used for off-line reference analysis of biomass, glucose, glycerol, acetate, and plasmid.

Reference Analyses. Biomass in units of dry cell weight (DCW) per volume of culture medium (g/L) was determined by centrifuging the cultivation samples, washing the pellet with 0.9% (w/v) NaCl, and drying at 80 °C until constant weight. The bacterial cell pellet and the supernatant obtained from sample centrifugation (Z160M centrifuge, Hermle Labortechnik, Germany) were frozen at −20 °C. Glucose, glycerol, and acetate were determined by HPLC with an L-6200 intelligent pump (Merck-Hitachi, UK), an L-7490 LaCromRi-detector (Merck), a D-2500 Chromato-integrator (Merck-Hitachi, Germany), and an Aminex fermentation monitor HPLC column (Bio-Rad, USA) maintained at 50 °C, using sulfuric acid at 0.6 mL/min as eluent. Plasmids were extracted from the bacterial cell by the alkaline cell lysis method, and subsequent plasmid concentration and purity degree were determined by hydrophobic interaction HPLC, as described in Scholz et al.[39]

Mid-Infrared Spectroscopy. The cell pellet obtained from the centrifugation of each 1 mL culture broth was resuspended with 0.9% NaCl 0.9% (w/v), so that an equivalent optical density of 6.0 (at 600 nm) in all samples was achieved. Triplicates of 25 μL of this suspension were placed on IR-transparent zinc selenide microtiter plates with 96 wells (Bruker Optics, Germany) and subsequently dehydrated for 2.5 h in a vacuum desiccator (ME2, Vaccubrand, Germany). The MIR spectra were recorded in transmission mode with an HTS-XT associated to Vertex-70 spectrometer (Bruker Optics), using a spectral resolution of 4 cm^{-1} and 40 scans per sample in the spectral range between 4000 and 500 cm^{-1}. Each culture sample was analyzed in triplicate.

Near-Infrared Spectroscopy. Near-infrared spectra were obtained using an NIR transflection fiber optic probe IN-271P (Bruker Optics), with a path length of 2 mm, coupled to a Vertex-70 spectrometer (Bruker Optics). A reference atmospheric air spectrum was acquired before the probe insertion in the bioreactor. The fiber optic probe submerged in the bioreactor was then steam sterilized simultaneously with the cultivation medium. NIR spectra were collected every 2 min in the 12 500–5400 cm^{-1} range, consisting of 32 co-added scans with 8 cm^{-1} resolution. The scanner velocity was set to 20 kHz and the aperture setting defined was 6 mm.

The MIR and NIR systems were operated with a dry air purging system. The compressed air is from outside the laboratory, to minimize the interferences from carbon dioxide oscillations in the laboratory air.

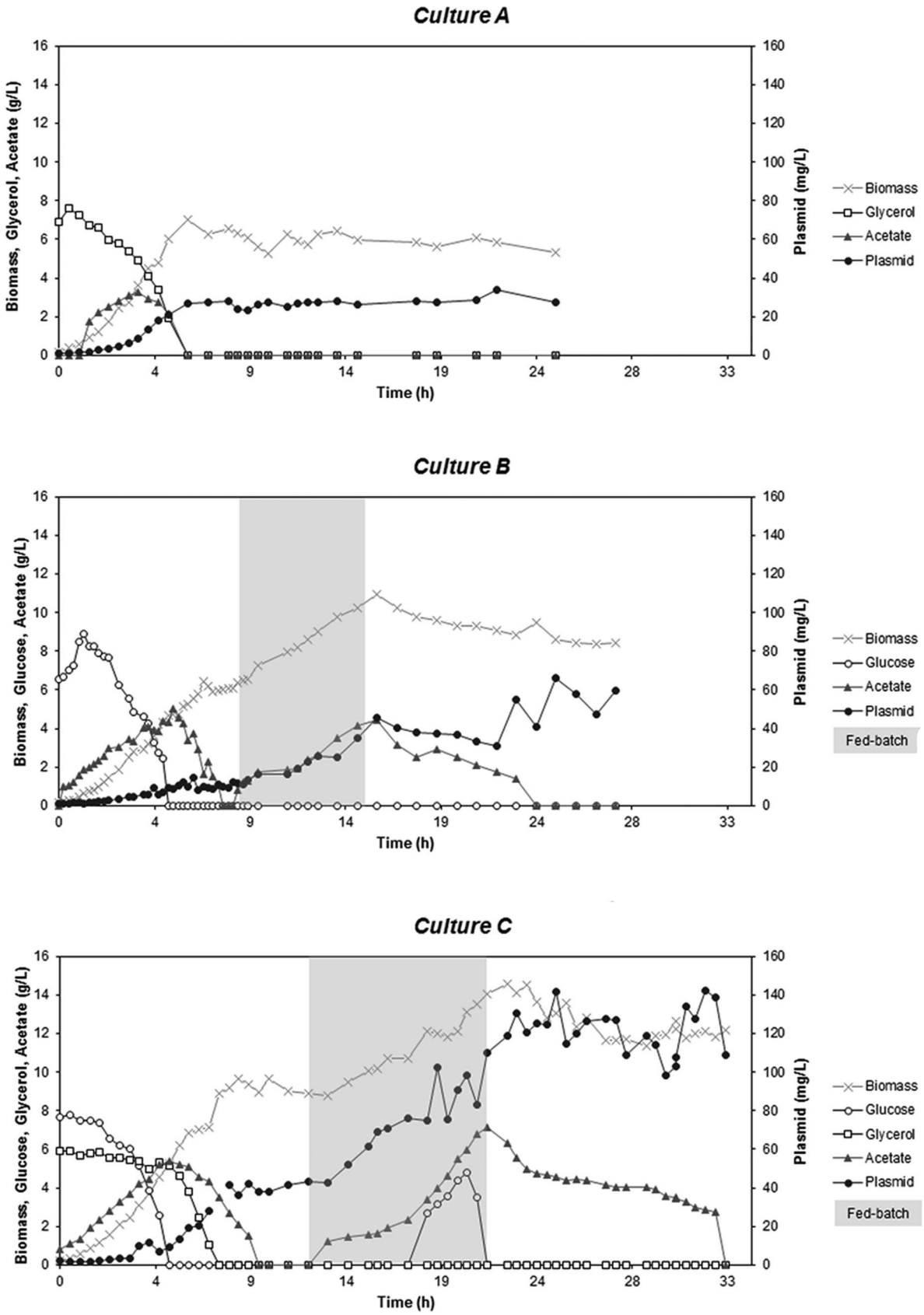

FIG. 1. Biomass, plasmid, glucose, glycerol, and acetate concentrations along the time of the three *E. coli* cultures (**A**, **B**, and **C**), conducted on glycerol (culture A), glucose (culture B), and glucose and glycerol (culture C) in the batch phase. On cultures B and C, after all acetate produced during the batch phase was consumed, an exponential feeding phase with glucose was started considering a $\mu = 0.18\ h^{-1}$, $Y_{X/S} = 0.6$ and S = 150 g/L. The feeding phase is represented in the graph by the gray area.

TABLE I. Glucose and glycerol concentrations used on three batch *E. coli* cultivations (cultures A, B, and C) and respective culture performance: maximum specific growth rate observed during each carbon source consumption and maximum acetate concentration achieved in the medium. The following parameters are relative to the time at which the maximum plasmid production was achieved: time, biomass, plasmid, plasmid-specific production, and plasmid productivity.

	Culture A	Culture B	Culture C
[Glucose] (g/L)	–	7.0	8.0
[Glycerol] (g/L)	7.0	–	6.0
Maximum [acetate] (g/L)	3.3	5	5.4
Time (h)	22	7	9.5
[Biomass] (g DCW/L]	5.9	5.6	9.4
[Plasmid] (mg/L)	34	14	42
Plasmid-specific production (mg/g DCW)	4.8	1.8	7.2
Plasmid productivity (mg/L/h)	1.5	2.1	4.4
Specific growth rate in glucose (h^{-1})	–	0.78	0.59
Specific growth rate in glycerol (h^{-1})	0.66	–	0.31

Chemometric Methods. The following chemometric analyses were conducted using OPUS version 7.2 software (Bruker, Germany).

Preprocessing. The following data preprocessing methods were evaluated: constant offset elimination, straight line subtraction, first and second derivatives, multiplicative scatter correction (MSC), standard normal variate (SNV), and different combinations of these methods. As derivatives usually broaden spectral noise, a Savitzky–Golay smoothing (using a five-point filter window and a second-order polynomial fit) was applied before the application of spectral derivatives.

Multivariate Data Analysis. The PLS method was used for predicting the concentration of the critical variables of the bioprocess from the spectral data as described previously.[30] The validation of the developed PLS models was performed using two different approaches: external validation, where a set of external samples not used for calibration was used for validating the model developed; and the leave-one-out cross-validation (LOOCV), where the calibration and validation are done by successively excluding a sample from the calibration set and using it as validation set, until all samples have been used for calibration and validation. The choice of the technique to be used was dependent on the number of samples available: for the biomass, plasmid, and acetate models, the root mean square error (RMSE) was calculated based on an independent test validation set due to the larger number of samples available; for the glucose and glycerol models, the model validation was obtained by LOOCV, as fewer samples were available.

TABLE II. Cultures performance obtained during the *E. coli* fed-batch cultures B and C, conducted with the same feeding strategy and using glucose, but with different batch phases (Table I): the maximum acetate concentration accumulated in the culture during the feeding phase; the remaining parameters are relative to time where the maximum plasmid production was achieved.

	Culture B	Culture C
Maximum [acetate] (g/L)	4.1	7.2
Time (h)	25	32.5
[Biomass] (g DCW/L]	8.6	12.1
[Plasmid] (mg/L)	66	142
Plasmid-specific production (mg/g DCW)	8.7	12.4
Plasmid productivity (mg/L/h)	2.6	4.4

Model Selection. Several PLS models were built based on the combination of the preprocessing techniques and wavelength selection as follows: the spectral region was divided into ten equal subregions, and the best combination of spectral regions and preprocessing techniques providing the best predictive model performance was chosen. The calculation started with one subregion, and after the best subregion had been determined the next subregions were subsequently added. This procedure was repeated for different combinations of the spectra preprocessing techniques.

The best PLS model was selected regarding a high coefficient of determination (R^2), a low predictive error (RMSE), and a low number of latent variables (LVs) covering enough data variance.

RESULTS AND DISCUSSION

To achieve robust PLS models to predict the concentrations of the critical variables of the bioprocesses from the IR spectra, PLS models were built based on cultures conducted with different media compositions regarding the carbon source and with different cultivation strategies (batch and fed-batch). The batch phases of cultivation A to C were conducted with glycerol (culture A), glucose (culture B), and a mixture of glucose and glycerol (culture C) (Fig. 1; Table I). After the batch phase, a feeding phase with glucose was started on cultures B and C (Fig. 1; Table II). The three cultures were monitored in high-throughput mode by MIR spectroscopy and in situ by NIR spectroscopy (Fig. 2).

Considering the batch phases of cultures A and B, culture A (conducted on glycerol) produced two times more plasmid than culture B (conducted on glucose) (Fig. 1; Table I), probably as a result of the lower specific growth rates and lower acetate productions associated with it.[16,39,47] Indeed, glycerol has previously been used as an alternative carbon source to glucose to minimize overflow metabolism, as a lower glycerol transport to the cell increases energetic metabolism efficiency, while reducing acetate production.[48,49] The minimization of the acetate production simultaneously increases the biomass yield and the product-specific production per biomass.[47,50] However, the lower specific growth rate associated to glycerol also results in a lower plasmid productivity. Therefore, the use of glycerol instead of

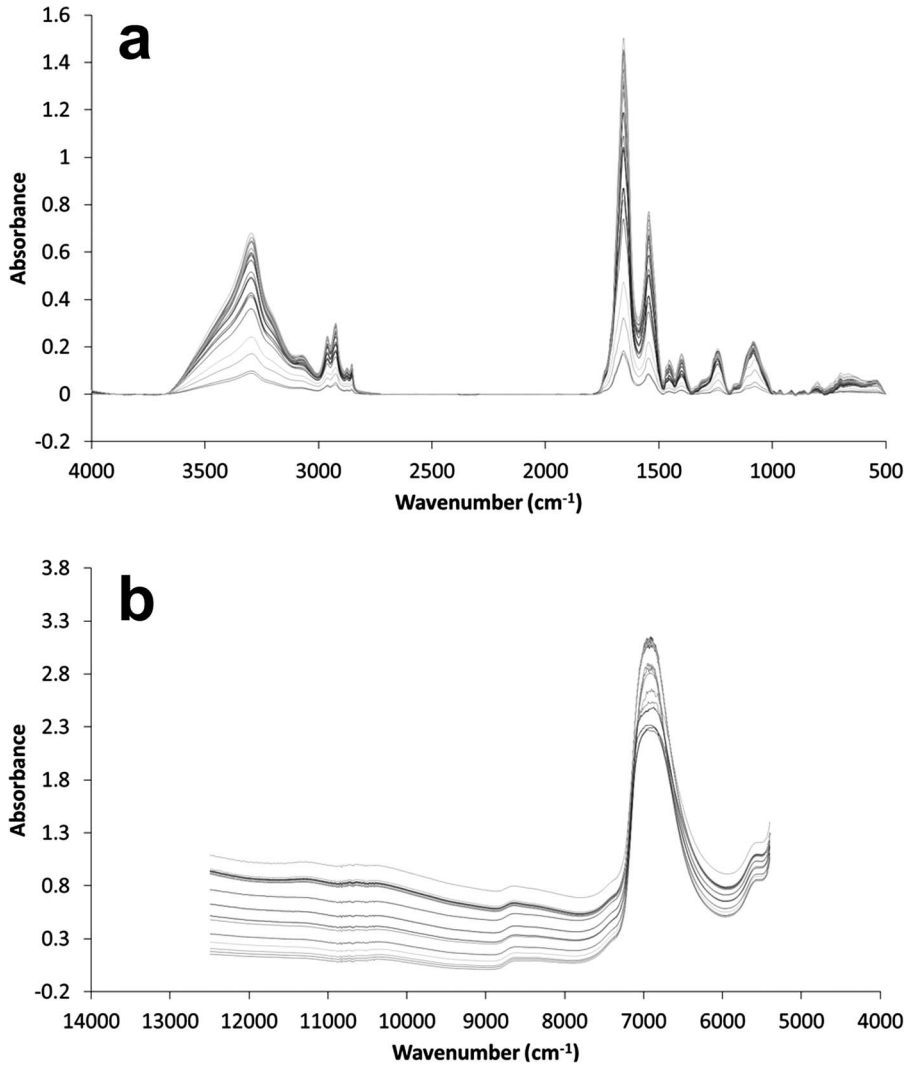

FIG. 2. MIR (**a**) and NIR (**b**) spectra acquired along the time of culture A.

glucose implies a lower specific growth rate, and consequently a higher plasmid-specific production per biomass, but at the expense of a lower plasmid productivity in relation to cultures based on glucose. Since the goal of the cultivation step is to simultaneously obtain maximum plasmid concentration, plasmid-specific production per biomass, and plasmid productivity, mixtures of glucose and glycerol as carbon sources are usually evaluated.[16,30,39,47] Indeed, culture C, conducted with a mixture of glucose and glycerol, presented the highest plasmid productivity (4.4 mg/L/h), plasmid concentration (42 mg/L), and plasmid-specific production

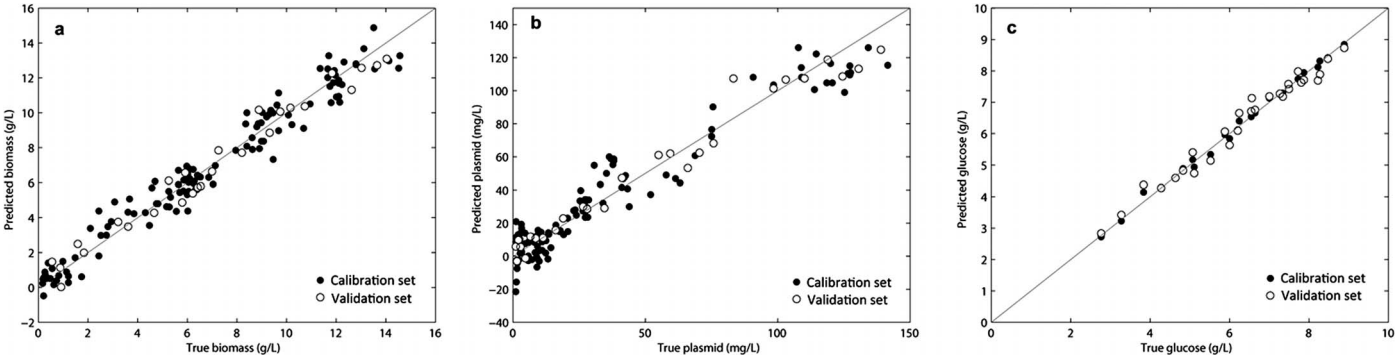

FIG. 3. Biomass, plasmid, and glucose concentrations along the time of the three cultures (**a**), (**b**), and (**c**), analyzed by conventional off-line assays in relation to the predictive values obtained by the PLS model based on at-line MIR spectroscopy.

TABLE III. Best PLS regression models obtained from high-throughput MIR spectroscopy for biomass, plasmid, glucose, glycerol, and acetate concentrations for cultures A, B, and C, concerning R^2; number of LVs and the RMSE; the preprocessing technique used; and the selected spectral regions, considering data (from LOOCV[a]).

	High-throughput monitoring by MIR spectroscopy							
	R^2	LV	RMSE	% of error	No. calibration samples	No. validation samples	Preprocessing	Spectral regions used (cm^{-1})
Biomass (g/L)	0.97	7	0.7	5	116	27	Second derivative	3300–2947 1200–500
Plasmid (mg/L)	0.97	8	9	6	116	27	First derivative and MSC	3300–2598 2249–1898
Glucose (g/L)[a]	0.97	8	0.3	3	28	–	First derivative	3998–3298 2949–2598 2249–1898
Glycerol (g/L)[a]	0.92	8	0.4	5	24	–	Second derivative	3998–3647 2949–1898 851–500
Acetate (g/L)	0.91	7	0.4	7	43	12	Straight line subtraction	2949–2598 2249–1898

per biomass (7.2 mg/g), in relation to the other two batch cultures (Table I).

Comparing the two fed-batch cultures, culture C produced about two times more plasmid and 1.6-fold higher plasmid productivity in relation to culture B (Fig. 1; Table II). This higher plasmid production and productivity resulted from the differences between the batch phases of the two cultures, since the feeding strategy was the same.

The data clearly show that slight differences in types and concentrations of carbon sources, as well as the cultivation strategy, have a relevant impact on the plasmid bioproduction, represented in the present study by a wide range of final biomass concentrations (between 5.6 and 12.1 g DCW/L), final plasmid concentrations (between 14 and 142 mg/L), final plasmid-specific productions per biomass (between 1.8 and 12.4 mg/g DCW), and final plasmid productivities (between 1.5 and 4.4 mg/L/h) (Tables I and II). Therefore, the plasmid bioproduction process represents a good example for which the development of monitoring techniques will increase the understanding of the bioprocess and consequently will promote the speed of the development process and the bioprocess control toward a reproducible process and a regulatory-compliant product.

Partial Least Squares Modeling of Mid-Infrared Spectra. Accurate PLS regression models were obtained for biomass, plasmid, and glucose concentrations, with an R^2 of 0.97 and an RMSE of 0.7 g/L, 9 mg/L, and 0.3 g/L, respectively, corresponding to percentage errors in relation to the range of units of the variables of 5, 6, and 3%, respectively (Fig. 3; Table III). Good PLS models were also obtained for glycerol and acetate, presenting an R^2 of 0.92 and 0.91, respectively, and an RMSE of 0.4 g/L (Table III). All models produced better estimates than those obtained by Scholz et al.,[39] when predicting the concentration of these compounds in five *E. coli* batch cultures with different initial medium compositions. These five cultures presented a distinct culture behavior, with maximum biomass concentrations between 6.7 and 12.8 g/L and maximum amounts of plasmid produced between 11 and 95 mg/L. Despite the large variability present in the cultures of the present work, a great improvement in model performance was achieved. This may be partially explained by the use of a larger number of samples used for PLS model building. It should also be noted that the prediction of the concentration of glucose, glycerol, and acetate from the MIR spectra of the cell pellets was only possible due to the correlations between the cell general metabolism and the concentration of these compounds in the extracellular medium. The PLS models obtained for all process variables were built as expected, based on some similar spectral regions, but also different combinations of spectral regions (Figs. 4a, 4b, 4c, 4d, and 4e; Table III), pointing out the specificity of each model.

The main absorbance regions of the MIR spectra reflect fundamental vibration modes of molecules, mainly from bond stretching and bond deformations. Other minor vibrational modes are twisting, wagging, rocking, and scissoring motions.[51,52] The main spectral regions for cells reflecting stretching vibrations include the following: 3600–2000 cm^{-1}, reflecting the majority of vibrations between X–H (where X is C, O, or N), present in amide A (at ~3200 cm^{-1}), and of CH_3 (at ~2955 and ~2870 cm^{-1}) and CH_2 (~2918 and 2850 cm^{-1}) groups from fatty acids; 1800–1500 cm^{-1}, reflecting mainly vibrations of double bonds (e.g., C=O, C=C, and C=N), present in amide I (~1695, 1685, and 1675 cm^{-1}) and amide II (~1550–1520 cm^{-1}) of proteins, of >C=O in phospholipids esters (~1740 cm^{-1}) or from carboxylic acids (~1715 cm^{-1}); and 1500–400 cm^{-1}, showing a variety of overlapped vibrations due to proteins, glucids, lipids, and nucleic acids, designated as fingerprinting region.[51,52]

The best PLS model was obtained by finding the best combination of spectral regions and preprocessing techniques providing the best predictive model performance, concerning a high R^2, a low RMSE, and a low number of LVs covering enough data variance. Spectral preprocessing allows removal of baseline offsets and enhancement of spectral information (e.g., spectral derivatives), as well as minimization of scatter effects due to particles of different sizes and shapes from the dehydrated cell pellets (e.g., MSC and SNV), and wavelength selection allows discarding of spectral

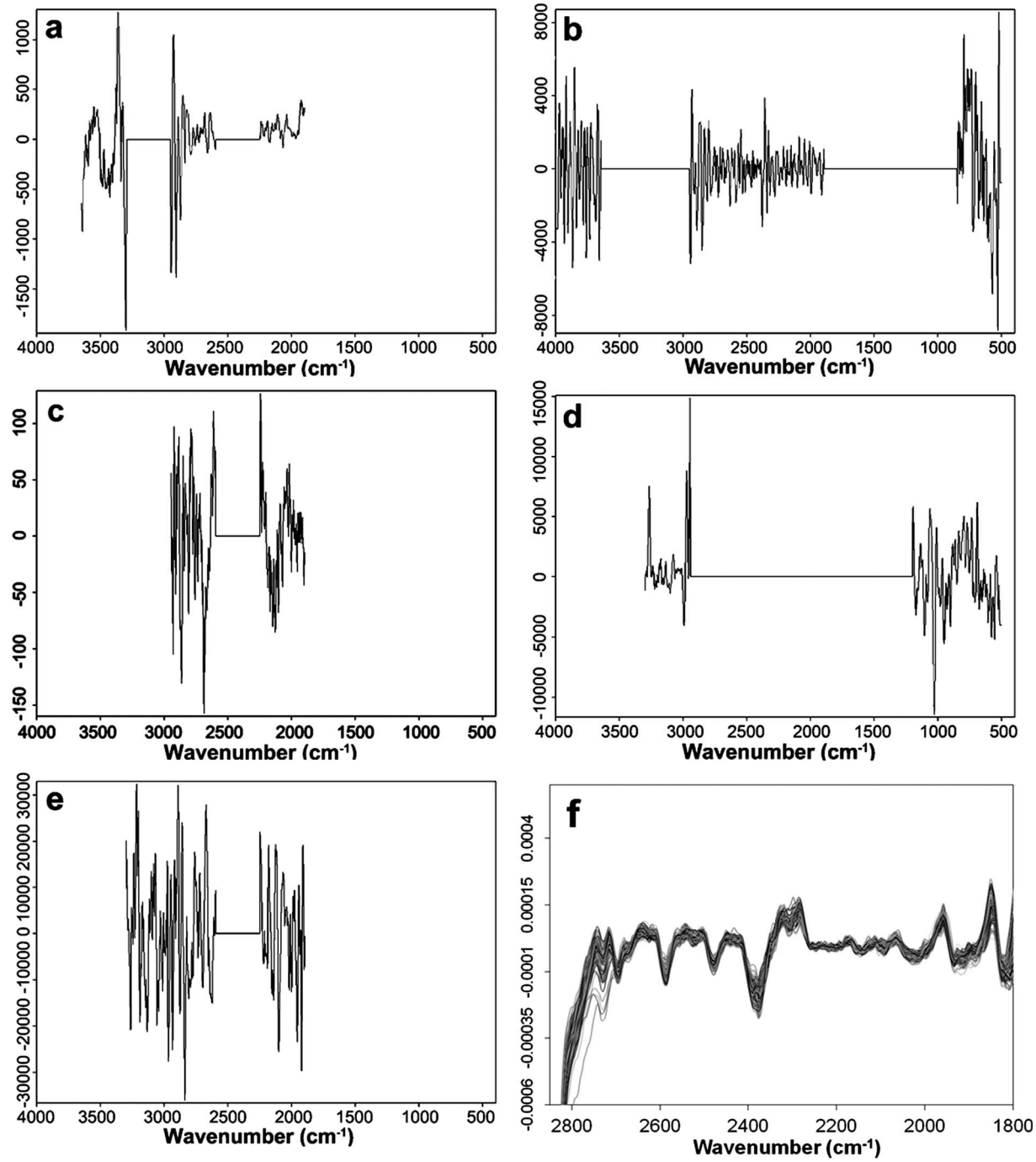

FIG. 4. PLS regression vectors obtained from the MIR models for (**a**) biomass, (**b**) plasmid, (**c**) glucose, (**d**) glycerol, and (**e**) acetate concentrations. (**f**) First derivative of MIR spectral data collected during culture A, highlighting the region between 1800 and 2800 cm^{-1}.

regions containing irrelevant information regarding the variables under study. For selecting the spectral subregions, the whole spectral region (from 4000 to 500 cm^{-1}) was divided into ten equal subregions that consequently resulted in PLS models built with spectral regions in multiples of 350 cm^{-1} (Fig. 4; Table III). The biomass model included the fingerprint region and the amide A and CH_3 stretching from fatty acids regions, in accordance to the contribution of the overall molecular composition for the biomass prediction (Fig. 4a). Interestingly, the plasmid model did not include the fingerprint region, but it did include the amide A, and CH_3 and CH_2 from fatty acids regions (Fig. 4b).

The CH_3 and CH_2 groups at these spectral regions are known to characterize the lipids from the cytoplasmatic membrane,[53–55] where an altered CH_3/CH_2 ratio has been associated to cell replication.[56] As plasmid replication is associated with cell growth,[39,45] the selection of these spectral regions probably results from the plasmid production with the cell division. The spectral region 2249–1898 cm^{-1} was also used for the construction of the plasmid model. Although no direct plasmid-related functional groups are known to absorb

TABLE IV. Best PLS regression models obtained from in situ NIR spectroscopy for biomass, plasmid, glucose, glycerol, and acetate concentrations for cultures A, B, and C, concerning R^2, number of LVs and RMSE, the preprocessing technique used, and the selected spectral regions, considering data (from LOOCV[a]).

	In situ monitoring by NIR spectroscopy							
	R^2	LV	RMSE	% of error	No. calibration samples	No. validation samples	Preprocessing	Spectral regions used (cm^{-1})
Biomass (g/L)	0.99	7	0.4	3	116	27	Constant offset elimination	11 078–9654 8949–7525
Plasmid (mg/L)	0.96	7	8	6	116	27	SNV	10 368–8945 6110–5400
Glucose (g/L)[a]	0.98	6	0.3	3	27	–	None	10 368–8235 6819–5400
Glycerol (g/L)[a]	0.96	7	0.2	3	23	–	None	8239–7526 6110–5400
Acetate (g/L)	0.90	4	0.4	7	44	15	Constant offset elimination	8949–7525

in this last region, some relevant information might be present, as highlighted in the regression vector of the PLS model (Fig. 4b) and in the first-derivative spectra of all culture samples (Fig. 4f), that explains the less accurate plasmid model when this region is removed (data not shown). This result is in accordance with a previous report on the prediction of plasmid by MIR spectral analysis of recombinant *E. coli* cell pellets,[39] where a region enclosing the subregion 1898–2025 cm^{-1} was found to be highly relevant for the plasmid prediction, as it was incorporated in more than 75% of the models explored by a wavenumber selection based on a Monte Carlo strategy. That the region considered in the present work (1898–2248 cm^{-1}) is larger than the region found by Scholz et al.[39] results from their use of spectral intervals that were approximately half the size of the intervals used here.

The glucose and glycerol models accounted for the region of OH vibrations (~3500 cm^{-1}) as these molecules are rich in hydroxyl groups (Figs. 4c and 4d). The glucose model also accounted for the less informative region between 2249 and 1898 cm^{-1}, and a small region characterized by the CH_3 and CH_2 vibrations between 2949 and 2598 cm^{-1}, whereas the glycerol model specifically included a small spectral window of the fingerprint region between 851 and 500 cm^{-1} and a large region between 2949 and 1898 cm^{-1}. The acetate model was the only model not requiring spectral derivatives, and it accounted for two of the three spectral regions used by the glucose model (Fig. 4e). The acetate model did not use the region characterized by the hydroxyl group.

Partial Least Squares Modeling of Near-Infrared Spectra. Good PLS models based on in situ NIR spectra were also achieved for all variables studied (Table IV). Highly accurate PLS regression models were achieved for biomass and glucose, with an R^2 of 0.98 and a low RMSE of 0.4 and 0.3 g/L, respectively (Fig. 5; Table IV). The biomass model yielded a high R^2 compared to previous reports on *E. coli* cultures, with lower prediction errors.[23,57] Accurate PLS models were also obtained for plasmid (Fig. 5; Table IV), yielding an R^2 of 0.96 and an RMSE of 8 mg/L, representing 3 and 6% of error, respectively, considering the range of concentrations of each variable used for model building.

The high RMSE associated to the plasmid model (6%) in relation to, e.g., the biomass model (3%), is in accordance to previous results by Scholz et al.[39] and Lopes et al.,[30] which reflects the high error associated to the reference method used to quantify the plasmid that ranges between 5 and 8% depending on the plasmid extraction yield from the cells. In the reference method, plasmid is extracted from the cells by alkaline lyses and subsequently quantified by HPLC. However, as the *E. coli* culture cells may present different metabolic states (e.g., content in nucleases), the plasmid extraction yield will depend of the cell growth phase and cultivation conditions. The RMSE of 6% associated to the plasmid model is thus in the range of error of the reference method.

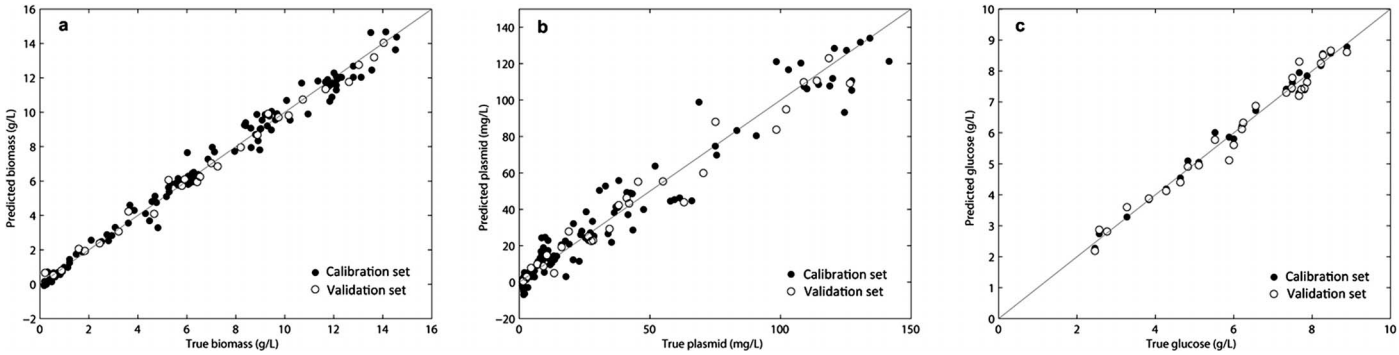

FIG. 5. Biomass, plasmid, and glucose concentrations along the time of the three cultures (**a**), (**b**), and (**c**), analyzed by conventional off-line assays in relation to the predictive values obtained by the PLS model based on in situ NIR spectroscopy.

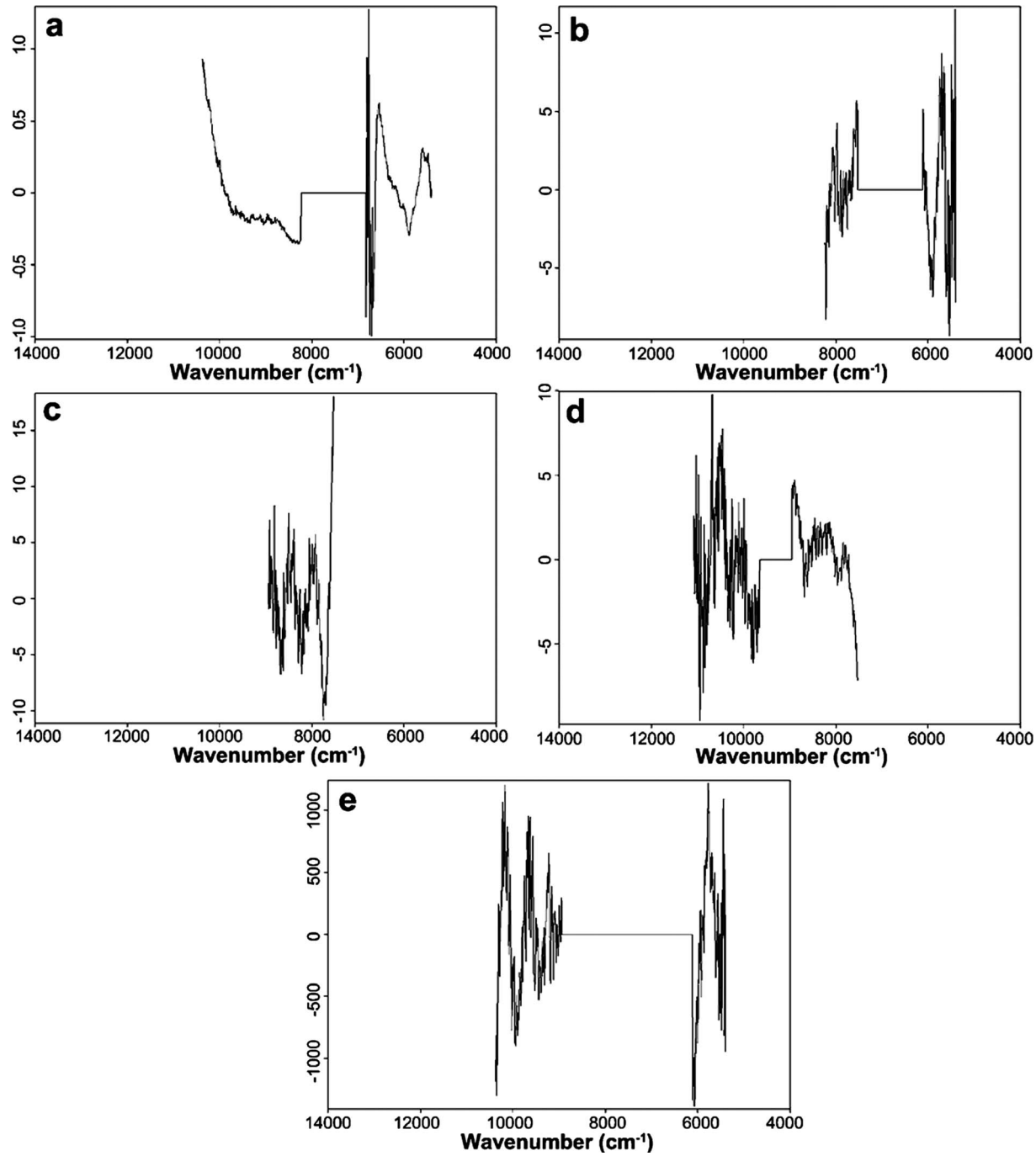

FIG. 6. PLS regression vectors obtained from the NIR models for (**a**) biomass, (**b**) plasmid, (**c**) glucose, (**d**) glycerol, and (**e**) acetate concentrations.

In the case of glycerol, an accurate PLS model with an R^2 of 0.96 and an RMSE of 0.2 g/L was obtained (Table IV). The acetate model presented the lowest R^2 (0.90), compared to the other models ($R^2 \geq 0.96$), and an RMSE of 0.4 g/L that corresponded to a high percentage error (8%) (Table IV). This less accurate PLS model can be related to the distinct metabolic behavior in cultures A to C, hence contributing to a greater complexity and consequently influencing the acetate prediction.

The PLS models for all process variables, based on NIR data, presented as expected some overlapping regions, but also different combinations of spectral regions (Fig. 6), thus contributing to model specificity. Due to the method used for selecting the optimum combination of spectral subregions, where the overall region (12 500–2500 cm^{-1}) was divided by ten subregions, the PLS models are based on subregions that are multiples of 710 cm^{-1} (Fig. 6; Table IV). The best biomass model accounted for spectral regions containing second overtones of CH bonds (CH, CH_2, and CH_3) vibrations, the third overtone of CH vibration, and second overtones of ROH and RNH_2 vibrations, highlighting the contribution of the cell overall molecular composition on the biomass model building (Fig. 6a). The best plasmid model presented contributions from the first overtone of CH, CH_2, and CH_3 vibrations and from the second overtone of

CH_3 and RNH_2 vibrations (Fig. 6b). The glucose model included almost all regions presented in the plasmid model, with exception for the second overtone of the CH_3 vibrations, and an additional region characteristic of the first overtone of RNH_2 vibrations (Fig. 6c). The best glycerol model accounted for the first overtone of CH, CH_2, and CH_3 vibrations, as the glucose model (Fig. 6d).

Mid-Infrared Versus Near-Infrared Models. Using exactly the same data from three batch cultures and two feeding phases, generally good PLS regression models in terms of number of LVs, R^2, and RMSE were achieved either with high-throughput MIR or in situ NIR data to predict all critical variables of the bioprocess: biomass, plasmid, glucose, glycerol, and acetate (Tables III and IV). Comparing the NIR and MIR setup, almost all variables presented very similar models concerning the R^2 and RMSE, with exceptions for biomass and glycerol. The biomass model based on NIR data presented 43% lower RMSE, similar R^2, and used the same number of LVs in relation to the MIR-based model. The glycerol model based on NIR data presented a 50% lower RMSE and a higher R^2 (of 0.96 in relation to 0.92) compared to the MIR-based model. The NIR-based models also used, in general, a lower number of LVs; for instance, the acetate and glucose models used six and four LVs, whereas the MIR models used eight and seven LVs, respectively. The plasmid model based on NIR data was very similar to the model obtained from the MIR data, presenting very similar RMSE (8 and 9 mg/L, respectively) and R^2 (0.96 and 0.97, respectively), and using a similar number of LVs (seven and eight, respectively).

The better predictive performance by the NIR setup for biomass and glycerol might result from the NIR setup acquiring the information directly from the culture broth, whereas in the MIR setup only the cell pellet is being analyzed, after sample extraction from the bioreactor, sample centrifugation, and subsequent cell pellet dehydration. Sandor et al.[58] and Sivakesava et al.[59] compared PLS models based on NIR and MIR spectral data to estimate critical variables in cultures of microorganisms and mammalian cells, and they obtained lower predictive errors for biomass and glucose models based on the MIR setup in relation to the NIR setup. However, in both cases, an ATR probe was used, enabling the on-line analysis in the MIR setup as opposed to the fast off-line high-throughput MIR analysis conducted in the present work. The high RMSE associated to the plasmid models (8–9 mg/L corresponding to a 6% error), obtained either from the MIR or the NIR setup, reflects the high error associated to the reference method based on the plasmid extraction from cells by alkaline lyses. Another reason for achieving a very good biomass model using the NIR setup might result from the NIR transflectance probe used; this probe copes with the highly dynamic culture environment, concerning stirring speed (between 100 and 1100 rpm), airflow rate (between 1.0 and 1.5 vvm), and cell concentrations (between 0.2 and 12 g DCW/L). Furthermore, the PLS models built on NIR data did not require the use of derivatives as preprocessing, as reported for other studies using in situ NIR probes.[32–34,36–38] For example, for glucose and glycerol models, no preprocessing was necessary, and only a constant offset elimination was applied for the biomass and acetate models.

CONCLUSIONS

This work shows the possibility of predicting all critical variables along highly dynamic recombinant *E. coli* cultures by fast off-line high-throughput MIR and in situ NIR spectroscopies. The choice of the technique, however, shall be determined in relation to the final purpose. In the case of control of the bioprocess, particularly at a large scale, the in situ analysis with the steam-sterilized NIR transflectance fiber optic probe should be used, as it enables the on-line monitoring in real time without the risk of culture contamination. It also provides better predictive models for some of the critical variables, e.g., biomass and glycerol. In case of optimization protocols conducted at a laboratory scale, for which multi-microbioreactors are used and costs or space constraints arise from the inclusion of NIR probes, high-throughput MIR spectroscopy by using plates with multi-microwells can represent a relevant alternative, as they enable the analysis of hundred samples from several cultures in a rapid and automatic mode.

ACKNOWLEDGMENTS

This work was supported by the PTDC/BIO/69242/2006 project from the Portuguese Foundation for Science and Technology (FCT). The authors gratefully acknowledge the postdoctoral scholarship from FCT (SFRH/BPD/73758/2010) and Prof. António Mendonça (Universidade da Beira Interior, Portugal) for technical support. This work was conducted in the BioEngineering Laboratory resulted from the protocol between Universidade Católica Portuguesa and the Instituto Politécnico de Lisboa.

1. R.A. Rader. ''FDA Biopharmaceutical Product Approvals and Trends in 2012: Up from 2011, but Innovation and Impact Are Limited''. BioProcess Int. 2013. 11(3): 18-27.
2. P.J. Declerck. ''Biologicals and Biosimilars: A Review of the Science and its Implications''. GaBi J. 2012. 1(1): 13-16.
3. R.J. Falconer, D. Jackson-Matthews, S.M. Mahler. ''Analytical Strategies for Assessing Comparability of Biosimilars''. J. Chem. Technol. Biotechnol. 2011. 86(7): 915-922.
4. C. Hakemeyer, U. Strauss, S. Werz, G.E. Jose, F. Folque, J.C. Menezes. ''At-Line NIR Spectroscopy as Effective PAT Monitoring Technique in Mab Cultivations During Process Development and Manufacturing''. Talanta. 2012. 90: 12-21.
5. S. Gnoth, M. Jenzsch, R. Simutis, A. Lübbert. ''Process Analytical Technology (PAT): Batch-to-Batch Reproducibility of Fermentation Processes by Robust Process Operational Design and Control''. J. Biotechnol. 2007. 132(2): 180-186.
6. European Medicines Agency (EMA). ''EMA-FDA Pilot Program for Parallel Assessment of Quality by Design Applications''. 2011. http://www.ema.europa.eu/docs/en_GB/document_library/Other/2011/03/WC500103621.pdf [accessed Jul 11 2014].
7. International Conference on Harmonization (ICH). Document ICH Q8–Q10. 2005–2010. www.ich.org [accessed Jul 11 2014].
8. United States Federal Food and Drug Administration (FDA). ''Guidance for Industry PAT–A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance''. 2004. http://www.fda.gov/downloads/Drugs/Guidances/ucm070305.pdf [accessed Feb 20 2014].
9. M. von Stosch, S. Davy, K. Francois, V. Galvanauskas, J.-M. Hamelink, A. Luebbert, M. Mayer, R. Oliveira, R. O'Kennedy, P. Rice, J. Glassey. ''Hybrid Modeling for Quality by Design and PAT–Benefits and Challenges of Applications in Biopharmaceutical Industry''. Biotechnol. J. 2014. 9(6): 719-726.
10. A.P. Teixeira, R. Oliveira, P.M. Alves, M.J.T. Carrondo. ''Advances in On-Line Monitoring and Control of Mammalian Cell Cultures: Supporting the PAT Initiative''. Biotechnol. Adv. 2009. 27(6): 726-732.

11. N.D. Lourenço, J.A. Lopes, C.F. Almeida, M.C. Sarraguça, H.M. Pinheiro. "Bioreactor Monitoring with Spectroscopy and Chemometrics: A Review". Anal. Bioanal. Chem. 2012. 404(4): 1211-1237.
12. D. Landgrebe, C. Haake, T. Höpfner, S. Beutel, B. Hitzmann, T. Scheper, M. Rhiel, K.F. Reardon. "On-Line Infrared Spectroscopy for Bioprocess Monitoring". Appl. Microbiol. Biotechnol. 2010. 88(1): 11-22.
13. V. Vojinovic, J.M.S. Cabral, L.P. Fonseca. "Real-Time Bioprocess Monitoring Part I: In Situ Sensors". Sens. Actuators B. 2006. 114(2): 1083-1091.
14. P. Harms, Y. Kostov, G. Rao. "Bioprocess Monitoring". Curr. Opin. Biotechnol. 2002. 13(2): 124-127.
15. P. Sampaio, L. Sousa, C.R.C. Calado, M.S. Pais, L.P. Fonseca. "Use of Chemometrics in the Selection of a *Saccharomyces cerevisiae* Expression System for Recombinant Cyprosin B Production". Biotechnol. Lett. 2011. 33(11): 2111-2119.
16. T. Silva, P. Lima, M. Roxo-Rosa, S. Hageman, L.P. Fonseca, 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. 2009. 23(4): 419-427.
17. C.R.C. Calado, C. Almeida, J.M.S. Cabral, L.P. Fonseca. "Development of a Fed-Batch Cultivation Strategy for the Enhance Production and Secretion of Cutinase by a Recombinant *Saccharomyces cerevisiae* SU50 Strain". J. Biosci. Bioeng. 2003. 96(2): 141-148.
18. R. Legmann, H.B. Schreyer, R.G. Combs, E.L. McCormick, A.P. Russo, S.T. Rodgers. "A Predictive High-Throughput Scale-Down Model of Monoclonal Antibody Production in CHO Cells". Biotechnol. Bioeng. 2009. 104(6): 1107-1120.
19. A. Buchenauer, M.C. Hofmann, M. Funke, J. Büchs, W. Mokwa, U. Schnakenberg. "Micro-Bioreactors for Fed-Batch Fermentations with Integrated Online Monitoring and Microfluidic Devices". Biosens. Bioelectron. 2009. 24(5): 1411-1416.
20. K. Isett, H. George, W. Herber, A. Amanullah. "Twenty-Four-Well Plate Miniature Bioreactor High-Throughput System: Assessment for Microbial Cultivations". Biotechnol. Bioeng. 2007. 98(5): 1017-1028.
21. Z. Zhang, G. Perozziello, P. Boccazzi, A.J. Sinskey, O. Geschke, K.F. Jensen. "Microbioreactors for Bioprocess Development". J. Lab. Autom. 2007. 12(3): 143-151.
22. M.A. Druy. "Applications for Mid-IR Spectroscopy in the Pharmaceutical Process Environment". Spectroscopy. 2004. 19(2): 61-63.
23. P. Roychoudhury, L.M. Harvey, B. McNeil. "The Potential of Mid Infrared Spectroscopy (MIRS) for Real Time Bioprocess Monitoring". Anal. Chim. Acta. 2006. 571(2): 159-166.
24. H. Kornmann, S. Valentinotti, P. Duboc, I. Marison, U. von Stockar. "Monitoring and Control of *Gluconacetobacter xylinus* Fed-Batch Cultures Using in situ Mid-IR Spectroscopy". J. Biotechnology. 2004. 113(1-3): 231-245.
25. D. Duygu, T. Baykal, I. Açikgöz, K. Yildiz. "Fourier Transform Infrared (FT-IR) Spectroscopy for Biological Studies". Gazi Univ. J. Sci. 2009. 22(3): 117-121.
26. S.S. Kadam, E. van der Windt, P.J. Daudey, H.J.M. Kramer. "A Comparative Study of ATR-FTIR and FT-NIR Spectroscopy for In Situ Concentration Monitoring During Batch Cooling Crystallyzation Processes". Cryst. Growth Des. 2010. 10(6): 2629-2640.
27. G. Vaccari, E. Dosi, A.L. Campi, G. Mantovani, Y.R.A. González-Vara, D. Matteuzi. "A Near Infrared Spectroscopy Technique for the Control of Fermentation Processes: An Application to Lactic Acid Fermentation". Biotechnol. Bioeng. 1994. 43(10): 913-917.
28. A.G. Cavinato, D.M. Mayes, Z. Ge, J. Callis. "Noninvasive Method for Monitoring Ethanol in Fermentation Processes Using Fiber-Optic Near-Infrared Spectroscopy". Anal. Chem. 1990. 62(18): 1977-1982.
29. Z. Ge, A.G. Cavinato, J.B. Callis. "Noninvasive Spectroscopy for Monitoring Cell Density in a Fermentation Process". Anal. Chem. 1994. 66(8): 1354-1362.
30. M.B. Lopes, G.A.L. Gonçalves, D. Felício-Silva, K.L.J. Prather, G.A. Monteiro, D.M.F. Prazeres, C.R.C. Calado. "In Situ NIR Spectroscopy Monitoring of Plasmid Production Processes: Effect of Producing Strain, Medium Composition and the Cultivation Strategy". J. Chem. Tech. Biotechnol. 2015. 90(2): 255-261.
31. P.N. Sampaio, K.C. Sales, F.O. Rosa, M.B. Lopes, C.R.C. Calado. "In Situ Near Infrared Spectroscopy Monitoring of Cyprosin Production by Recombinant *Saccharomyces cerevisiae* Strain". J. Biotechnol. 2014. 188C: 148-157.
32. M. Navrátil, A. Norberg, L. Lembrén L, 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. 2005. 115(1): 67-79.
33. S. Tosi, M. Rossi, E. Tamburini, G. Vaccari, A. Amaretti, D. Matteuzzi. "Assessment of In-Line Near-Infrared Spectroscopy for Continuous Monitoring of Fermentation Processes". Biotechnol. Prog. 2003. 19(6): 1816-1821.
34. E. Tamburini, G. Vaccari, S. Tosi, A. Trilli. "Near-Infrared Spectroscopy: A Tool for Monitoring Submerged Fermentation Process Using an Immersion Optical-Fiber Probe". Appl. Spectrosc. 2003. 57(2): 132-138.
35. S.A. Arnold, J. Crowley, N. Woods, L.M. Harvey, B. McNeil. "In Situ Near Infrared Spectroscopy to Monitor Key Analytes in Mammalian Cell Cultivation". Biotechnol. Bioeng. 2003. 84(1): 13-19.
36. S.A. Arnold, R. Gaensakoo, L.M. Harvey, 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. 2002. 80(4): 405-413.
37. C. Cimander, C.F. Mandenius. "Online Monitoring of a Bioprocess Based on a Multi-Analyser System and Multivariate Statistical Process Modeling". J. Chem. Technol. Biotechnol. 2002. 77(10): 1157-1168.
38. J. Schenk, I.W. Marison, U. von Stockar. "Simplified Fourier-Transform Mid-Infrared Spectroscopy Calibration Based on a Spectra Library for the On-Line Monitoring of Bioprocesses". Anal. Chim. Acta. 2007. 591(1): 132-140.
39. T. Scholz, V.V. Lopes, C.C. Calado. "High-Throughput Analysis of the Plasmid Bioproduction Process in *Escherichia coli* by FTIR Spectroscopy". Biotechnol. Bioeng. 2012. 109(9): 2279-2285.
40. S. Gross-Selbeck, G. Margreiter, C. Obinger, K. Bayer. "Fast Quantification of Recombinant Protein Inclusion Bodies Within Intact Cells by FT-IR Spectroscopy". Biotechnol. Prog. 2007. 23(3): 762-766.
41. H. Büning. "Gene Therapy Enters the Pharmamarket: The Short Story of a Long Journey". EMBO Mol. Med. 2013. 5(1): 1-3.
42. A. Shibui, S. Nakae, J. Watanabe, Y. Sato, M.E.M. Tolba, J. Doi, T. Shiibashi, S. Nogami, S. Sugano, N. Hozumi. "Screening of Novel Malaria DNA Vaccine Candidates Using Full-Length cDNA Library". Exp. Parasitol. 2013. 135(3): 546-550.
43. S.A. Kalams, S.D. Parker, M. Elizaga, B. Metch, S. Edupuganti, J. Hural, S. De Rosa, D.K. Carter, K. Rybczyk, I. Frank, J. Fuchs, B. Koblin, D.H. Kim, P. Joseph, M.C. Keefer, L.R. Baden, J. Eldridge, J. Boyer, A. Sherwat, M. Cardinali, M. Allen, M. Pensiero, C. Butler, A.S. Khan, J. Yan, N.Y. Sardesai, J.G. Kublin, D.B. Weiner. "Safety and Comparative Immunogenicity of an HIV-1 DNA Vaccine in Combination with Plasmid Interleukin 12 and Impact of Intramuscular Electroporation for Delivery". J. Infect. Dis. 2013. 208(5): 818-829.
44. C. Coban, K. Kobiyama, T. Aoshi, F. Takeshita, T. Horii, S. Akira, K.J. Ishii. "Novel Strategies to Improve DNA Vaccine Immunogenicity". Curr. Gene Ther. 2011. 11(6): 479-484.
45. K.J. Prather, S. Sagar, J. Murphy, M. Chartrain. "Industrial Scale Production of Plasmid DNA for Vaccine and Gene Therapy: Plasmid Design, Production, and Purification". Enzyme Microb. Technol. 2003. 33(7): 865-883.
46. C.R.C. Calado, M. Mannesse, M. Egmond, J.M.S. Cabral, L.P. Fonseca. "Production of Wild Type and Peptide Fusion Cutinases by Recombinant *Saccharomyces cerevisiae*". Biotechnol. Bioeng. 2002. 78(6): 692-698.
47. M.B. Lopes, G. Martins, C.R.C. Calado. "Kinetic Modeling of Plasmid Bioproduction in Escherichia coli Cultures Conducted in Different Media Compositions". J. Biotechnol. 2014. 186: 38-48.
48. D.J. Korz, U. Rinas, K. Hellmuth, E.A. Sanders, W.D. Deckwer. "Simple Fed-Batch Technique for High Cell Density Cultivation of *Escherichia coli*". J. Biotechnol. 1995. 39(1): 59-65.
49. R. Hansen, N.T. Eriksen. "Activity of Recombinant GST in *Escherichia coli* Grown on Glucose and Glycerol". Process Biochem. 2007. 42(8): 1259-1263.
50. J. Trevisan, P.P. Angelov, P.L. Carmichael, A.D. Scott, F.L. Martin. "Extracting Biological Information with Computational Analysis of Fourier-Transformed Infrared (FTIR) Biospectroscopy Datasets: Current Practices to Future Perspectives". Analyst. 2012. 137(14): 3202-3215.
51. G. Bellisola, C. Sorio. "Infrared Spectroscopy and Microscopy in Cancer Research and Diagnosis". Am. J. Cancer Res. 2012. 2(1): 1-21.

52. K. Maquelin, C. Kirschner, L.-P. Choo-Smith, N. van den Braak, H.P. Endtz, D. Naumann, G.J. Puppels. ‘‘Identification of Medically Relevant Microorganisms by Vibrational Spectroscopy’’. J. Microbiol. Methods 2002. 51(3): 255-271.
53. C. Paluszkiewicz, W.M. Kwiatek. ‘‘Analysis of Human Cancer Prostate Tissues Using FTIR Microspectroscopy and SRIXE Techniques’’. J. Mol. Struct. 2001. 565-566: 329-334.
54. E. Bogomolny, M. Huleihel, Y. Suproun, R.K. Sahu, S. Mordechai. ‘‘Early Spectral Changes of Cellular Malignant Transformation Using Fourier transform Infrared Microspectroscopy’’. J. Biomed. Opt. 2007. 12(2): 024003.
55. M. Huleihel, A. Salman, V. Erukhimovitch, J. Ramesh, Z. Hammody, S. Mordechai. ‘‘Novel Spectral Method for the Study of Viral Carcinogenesis In Vitro’’. J. Biochem. Biophys. Methods. 2002. 50(2-3): 111-121.
56. N. Jamin, L. Miller, J. Moncuit, W.H. Fridman, P. Dumas, J.L. Teillaud. ‘‘Chemical Heterogeneity in Cell Death: Combined Synchrotron IR and Fluorescence Microscopy Studies of Single Apoptotic and Necrotic Cells’’. Biopolymers. 2003. 72(5): 366-373.
57. V. Di Egidio, N. Sinelli, G. Giovanelli, A. Moles, E. Casiraghi. ‘‘NIR and MIR Spectroscopy as Rapid Methods to Monitor Red Wine Fermentation’’. Eur. Food Res. Technol. 2010. 230(6): 947-955.
58. M. Sandor, F. Rüdinger, R. Bienert, C. Grimm, D. Solle, T. Scheper. ‘‘Comparative Study of Non-Invasive Monitoring Via Infrared Spectroscopy for Mammalian Study Cultivations’’. J. Biotechnol. 2013. 168(4): 636-645.
59. S. Sivakesava, J. Irudayaraj, D. Ali. ‘‘Simultaneous Determination of Multiple Components in Lactic Acid Fermentation Using FT-MIR, NIR and FT-Raman Spectroscopic Techniques’’. Process Biochem. 2001. 37(4): 371-378.