



Prediction of intracellular storage polymers using quantitative image analysis in enhanced biological phosphorus removal systems[☆]



Daniela P. Mesquita^a, Cristiano Leal^b, Jorge R. Cunha^a, Adrian Oehmen^c,
A. Luís Amaral^{a,b}, Maria A.M. Reis^c, Eugénio C. Ferreira^{a,*}

^a IBB–Institute for Biotechnology and Bioengineering, Centre of Biological Engineering, Universidade do Minho, Campus de Gualtar, 4710–057 Braga, Portugal

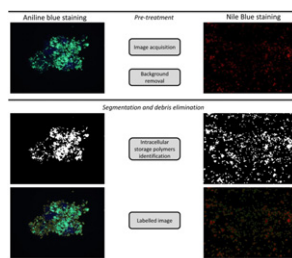
^b Instituto Superior de Engenharia de Coimbra, Instituto Politécnico de Coimbra, Rua Pedro Nunes, Quinta da Nora, 3030–199 Coimbra, Portugal

^c CQFB/REQUIMTE, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829–516 Caparica, Portugal

HIGHLIGHTS

- ▶ Fluorescence staining and image analysis were used to monitor an EBPR system.
- ▶ Intracellular storage polymers were measured in anaerobic and aerobic stages.
- ▶ Partial least squares (PLS) were used to model the polymers concentrations.
- ▶ Previous stage identification improved the assessment of the polymer concentrations.
- ▶ A novel method for intracellular storage polymers assessment was established.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 11 August 2012

Received in revised form 17 January 2013

Accepted 4 February 2013

Available online 16 February 2013

Keywords:

Partial least squares

EBPR wastewater treatment systems

Image analysis

Intracellular polymers

Polyhydroxyalkanoates

Glycogen

ABSTRACT

The present study focuses on predicting the concentration of intracellular storage polymers in enhanced biological phosphorus removal (EBPR) systems. For that purpose, quantitative image analysis techniques were developed for determining the intracellular concentrations of PHA (PHB and PHV) with Nile blue and glycogen with aniline blue staining. Partial least squares (PLS) were used to predict the standard analytical values of these polymers by the proposed methodology. Identification of the aerobic and anaerobic stages proved to be crucial for improving the assessment of PHA, PHB and PHV intracellular concentrations. Current Nile blue based methodology can be seen as a feasible starting point for further enhancement. Glycogen detection based on the developed aniline blue staining methodology combined with the image analysis data proved to be a promising technique, toward the elimination of the need for analytical off-line measurements.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Enhanced biological phosphorus removal (EBPR) is commonly employed to remove phosphorus (P) from wastewater treatment plants (WWTP). EBPR is performed by operating the system through sequentially alternating anaerobic and aerobic conditions [1]. This strategy was specifically developed to favor the

[☆] Paper presented at the XIII Conference on Chemometrics in Analytical Chemistry (CAC 2012), Budapest, Hungary, 25–29 June 2012.

* Corresponding author. Tel.: +351 253 604 407; fax: +351 253 678 986.

E-mail address: ecferreira@deb.uminho.pt (E.C. Ferreira).

polyphosphate accumulating organisms (PAOs) versus other heterotrophic organisms when competing for organic substrates (such as the case of volatile fatty acids, VFA). In the anaerobic stage, PAOs are able to take up VFA and store them as intracellular storage polymers, namely polyhydroxyalkanoates (PHA). According to the current knowledge of PAO metabolism, the energy for these anaerobic transformations is generated through hydrolysis of their internally stored polyphosphate (poly-P) and glycogen, while glycogen glycolysis also produces the reducing power necessary to maintain the cell redox balance [2,3]. This metabolism step results in orthophosphate release into the bulk. Under the presence of aerobic or anoxic conditions, PAOs oxidize the internal stored PHA for growth, glycogen replenishment and poly-P recovery accumulating P as poly-P inclusions. P removal is achieved by removing the excess sludge with a high poly-P level at the end of the aerobic stage [4].

Glycogen accumulating organisms (GAOs) are also capable of anaerobic VFA uptake in EBPR, however, without contributing to P removal [5]. Therefore, GAOs are a main cause of EBPR failure as they are selected under similar conditions. Several studies have been performed in order to comprehend strategies that influence the competition between PAOs and GAOs as well as the modeling of both types of bacterial activities [4,6–10]. Key parameters that have been identified include the carbon source, pH and temperature [10–12].

PHA (where the most common monomers found are polyhydroxybutyrate (PHB) and poly-hydroxyvalerate (PHV)) and glycogen are very important intracellular storage compounds in EBPR systems. PHA synthesized by microorganisms possesses a potentially broad range of commercial applications as biodegradable plastics, with thermoplastic characteristics of polypropylene and polyethylene. Conventionally, PHA production based on pure culture and carbon sources led to high costs. However high quantities of intracellular PHA can be produced from activated sludge and other mixed microbial cultures [13–15]. On the other hand, glycogen is a glucose based polysaccharide with a very important role in microbial communities under diverse environmental conditions, of high relevance in metabolic models for estimating kinetic and stoichiometric parameters of the biological process [15–18].

Monitoring these intracellular storage polymers in mixed cultures is usually performed through laborious and time consuming off-line chemical analysis including gas or high-performance liquid chromatography. Thus, there is a clear need to develop new and faster techniques to promptly monitor these processes. For the vast majority of the studies concerning intracellular storage polymers, no attempt to automatically extract quantitative information from microscopic images by image analysis has been made. Nevertheless, image analysis techniques have the potential to be a non-invasive and rapid means of assessing the amount of different storage polymers inside microbial cells, facilitating the evaluation of these important biotechnological processes. In fact, in activated sludge monitoring, several studies were already performed for the assessment of biomass morphological changes using bright-field or phase-contrast techniques [19–23] or even using fluorescent staining for gram or viable/damaged cells quantification [24,25]. The potential of quantitative image analysis was already demonstrated through the use of multivariate statistical analysis (principal component analysis – PCA and partial least squares – PLS) combined with operational parameters where distinct abnormalities were identified, and sludge settling ability was predicted [25,26].

This research focused on the automatic identification and quantification of PHA and glycogen intracellular concentrations in microbial structures, since off-line analysis is labor intensive and difficult to implement in full-scale plants. Samples were collected at the end of the anaerobic and aerobic stages for PHA (including PHB and PHV) and glycogen determination and for microscopy

inspection with staining procedures. All of the information was treated using multivariate statistical analysis. PLS was applied incorporating image analysis alongside off-line analysis. The potential of PLS was verified for the prediction of PHA and glycogen intracellular storage polymer compounds.

2. Materials and methods

2.1. Experimental setup

A lab-scale sequencing batch reactor (SBR) fed with synthetic wastewater containing volatile fatty acids (VFAs) and orthophosphate was used. The reactor had a working volume of 4 L and was operated with a cycle time of 6 h consisting of 120 min anaerobic stage, 180 min aerobic stage, 55 min settling/decanting stage, and 5 min feed. In each cycle, 2 L of synthetic wastewater (described below) was fed to the reactor in the first 5 min of the anaerobic period, resulting in a hydraulic retention time (HRT) of 12 h. At the end of the aerobic stage, 125 mL of excess sludge was removed, resulting in a sludge retention time (SRT) of 8 days. A stirrer speed of 500 rpm was maintained during each cycle, except for the settling period. Two on/off control valves were used during anaerobic and aerobic periods to control the gas flow. Nitrogen gas was bubbled into the reactor during the anaerobic period at a flow rate of approximately 0.5 L min^{-1} to maintain strict anaerobic conditions in the reactor. In the aerobic period, compressed air supply was used and the dissolved oxygen (DO) concentration was controlled around 3 mg L^{-1} . The pH was controlled during both the anaerobic and aerobic periods around 7 using two-way controller pumps that dosed 0.1 M HCl or 0.1 M NaOH when the pH was above/below the set point, and the temperature was controlled at 30°C .

2.2. Synthetic wastewater composition

The system was fed with synthetic wastewater containing acetate as the main carbon source. A phosphate solution (K_2HPO_4 and KH_2PO_4) was prepared separately. The chemical oxygen demand (COD) concentration in the feed was $200 \text{ mg COD L}^{-1}$, while the phosphorus concentration was $2 \text{ mg P-PO}_4 \text{ L}^{-1}$, which yielded a COD/P ratio of $100 \text{ mg COD mg P-PO}_4^{-1}$ in order to provide selective advantages to GAOs over PAOs. The concentration of the other nutrients added in the synthetic feed are listed below (per L) [15]: $0.59 \text{ g NH}_4\text{Cl}$, $0.95 \text{ g MgSO}_4 \cdot 7\text{H}_2\text{O}$, $0.44 \text{ g CaCl}_2 \cdot 2\text{H}_2\text{O}$, $11.66 \text{ mg allyl-N thiourea (ATU)}$, a nitrification inhibitor, 0.03 g EDTA , and 3.16 mL of a trace metals solution for biomass maintenance. The trace metals solution has been also described in Smolders et al. [15] and consisted of (per L): $1.5 \text{ g FeCl}_3 \cdot 6\text{H}_2\text{O}$, $0.15 \text{ g H}_3\text{BO}_3$, $0.03 \text{ g CuSO}_4 \cdot 5\text{H}_2\text{O}$, 0.18 g KI , $0.12 \text{ g MnCl}_2 \cdot 4\text{H}_2\text{O}$, $0.06 \text{ g Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$, $0.12 \text{ g ZnSO}_4 \cdot 7\text{H}_2\text{O}$, $0.15 \text{ g CoCl}_2 \cdot 6\text{H}_2\text{O}$.

2.3. Analytical procedures

Samples were collected at the end of the anaerobic and aerobic stages, for glycogen, and PHAs (including poly- β -hydroxybutyrate (PHB) and poly- β -hydroxyvalerate (PHV)) determination. The sludge obtained from each stage of the process was lyophilized during 48 h.

Glycogen was determined via the modified method of Bond et al. [27]. Pre-weighed samples of lyophilized sludge were added to 2 mL of 0.6 M HCl and digested for 2 h at 100°C . After cooling, the samples were centrifuged and 1 mL of the supernatant liquid was filtered through a $0.2 \mu\text{m}$ membrane filter and analyzed for glucose by high-performance liquid chromatography (HPLC). Chromatographic separation was performed using a Varian Metacarb 87H column ($300 \text{ mm} \times 7.8 \text{ mm}$, Varian, USA) under the following conditions: mobile phase $0.005 \text{ M H}_2\text{SO}_4$, flow rate 0.6 mL min^{-1} , and

Table 1
Sample numbers and concentration ranges for PHB, PHV, PHA, and glycogen at the end of anaerobic (AN) and aerobic (AE) stages.

Storage polymer	Number of samples	Polymer concentration at the end of AN stage (mg L ⁻¹)		Number of samples	Polymer concentration at the end of AE stage (mg L ⁻¹)		Overall concentration (mg L ⁻¹)	
		AN	Min		Max	AE	Min	Max
PHB	54	2.17	241.62	54	1.34	54.27	1.34	241.62
PHV	54	0.00	75.32	54	0.00	47.38	0.00	75.327
PHA	54	3.80	316.95	54	3.82	68.51	3.79	316.95
Glycogen	35	0.42	72.40	35	1.12	77.55	0.42	77.55

column temperature 60 °C. The system was comprised of a Jasco chromatograph 880-PU intelligent pump equipped with a 830-IR intelligent refraction-index detector and a AS-2057 Plus intelligent auto sampler (Jasco, Tokyo, Japan). The measured glucose equivalents were assumed to be derived from intracellular glycogen [4,11].

PHB and PHV content was measured by gas chromatography (GC) using the adapted method developed by Smolders et al. [15]. Pre-weighed samples of lyophilized sludge were transferred to glass tubes. The polymers were esterified with HCl:1-Propanol (1.5 mL) and extracted with dichloromethane (1.5 mL) including an internal standard (benzoic acid). The mixture was digested at 100 °C for 3.5 h. Quantification was made in a GC system (Varian 3800 instrument, Varian Inc., USA) equipped with a flame ionization detector. PHB and PHV were separated using a TRWAX capillary column (Teknokroma, Spain), with helium as the carrier gas at 1.0 mL min⁻¹. Temperatures of the split injection and detector were 220 °C and 250 °C, respectively. The initial oven temperature was 50 °C for 2 min, with a 10 °C min⁻¹ ramp to 225 °C and then maintained for 10 min. Data were analyzed using the acquisition and integration software Star Chromatography Workstation v. 6.30 (Varian Inc., USA) and calibration curves obtained for PHB and PHV.

The number of collected samples, as well as the concentration ranges for each of the studied polymers, is presented in Table 1.

2.4. Staining procedure

Fresh sludge samples were collected at the end of the anaerobic and aerobic stages and fixed in paraformaldehyde (PFA), then washed with phosphate buffer saline solution (PBS), and stored in a PBS/ethanol solution at -20 °C prior to further analysis [28,29].

Intracellular PHA was observed using Nile blue staining. Two comparative studies were performed using air dried smears on glass slides and biomass suspension (data not shown). Better results for quantitative image analysis were obtained with biomass suspension. The methodology used was adapted from Ostle and Holt [30]. 1 mL of fixed cell suspension was incubated at 55 °C for 10 min with 1 drop of Nile blue. The sample was then centrifuged at 4500 rpm for 5 min. Pellet cells were washed with 0.9% NaCl and centrifuged using the same conditions. The excess of stain solution was removed using 8% acetic acid for 1 min and centrifuged again. Pellet cells were then re-suspended in 0.9% NaCl. In order to guarantee good quality images for the implementation of quantitative image analysis, a tissue grinder was used to disrupt the granules formed during the staining procedure.

Intracellular glycogen storage was observed using Aniline blue staining. Although aniline blue is generally used to detect glycogen in tissue sections [31], in activated sludge the application of aniline blue staining procedures for glycogen detection has been questioned and not frequently used due to inconsistent results [32]. In order to try to overcome the limitations of the existing methodologies, two comparative studies (data not shown) were performed using 2% acid solution (acetic acid with pH at 3.5) and alternatively 2% alkaline solution (phosphate buffer with pH at 7.2) during 20 min reaction [31] with biomass suspension. Upon

comparison between the quantitative image analysis results, for the acquired fluorescent images of the two studies, and the sample glycogen analytical concentration obtained via HPLC, better results were obtained using the 2% alkaline solution [31]. This agreed with the findings of Shennawy et al. [31]. Thus, the methodology used for biomass suspension was as follows. 1 mL of fixed cells suspension was incubated for 20 min, with 1 drop of aniline blue prepared with 2% alkaline solution. The sample was then centrifuged at 4500 rpm for 5 min. The pellet cells were washed and suspended in 0.9% NaCl. As above, a tissue grinder was also used to disrupt the granules formed during the staining procedure.

2.5. Epifluorescence microscopy

Periodic samples were collected to assess the physiology of the activated sludge by staining and epifluorescence microscopy. A recalibrated micropipette with a sectioned tip at the end was used to deposit samples on slides. Three slides per sample were used, and for each slide a volume of 10 µL of the stained sludge samples was covered with a 20 mm × 20 mm cover slip, for visualization and image acquisition. Images were acquired in the upper, middle and bottom of the slide resulting in a total of 150 images (3 × 50 images per slide). The slides were examined by means of an epifluorescence microscope Olympus BX51 (Olympus, Tokyo, Japan) at 400× total magnification, and constant exposure times for intensity measurements. For Nile blue staining, a long pass filter was used with an excitation bandpass of 530–550 nm and emission cut off at 591 nm, with the stained PHA inclusions presenting emission spectra around 625–675 nm, for this excitation bandpass. For aniline blue staining, a long pass filter was used with an excitation bandpass of 365–370 nm and emission cut off at 421 nm, with the stained glycogen inclusions presenting emission spectra around 525–575 nm, for this excitation bandpass. Images were acquired at 1360 × 1024 pixels, and 24-bit RGB format (8 bit red, 8 bit green and 8 bit blue channels) through the commercial software CellB (Olympus, Tokyo, Japan).

2.6. Image analysis methodology

The image processing and analysis methodology regarding glycogen inclusions was based on the identification and quantification of both individual glycogen granules and glycogen inclusion regions. A schematic representation of the image analysis procedure is presented in Fig. 1. The first step of the image processing program, regarding glycogen inclusions determination, converted the RGB image into the L*a*b* color space. In order to increase the distinction from the greenish glycogen coloration (aniline blue stained glycogen inclusions emitted at a green spectra wavelength of 525–575 nm) from the remaining blue activated sludge, the b* channel was chosen for subsequent treatment. Nonetheless, the first step of this program consisted on subtracting the a* channel from the b* channel, followed by an automated (Otsu method) segmentation. The resulting binary image was considered as the glycogen inclusion region. Regarding individual glycogen granules identification, the first step of the image processing program also consisted on converting the RGB image into the L*a*b* color space.

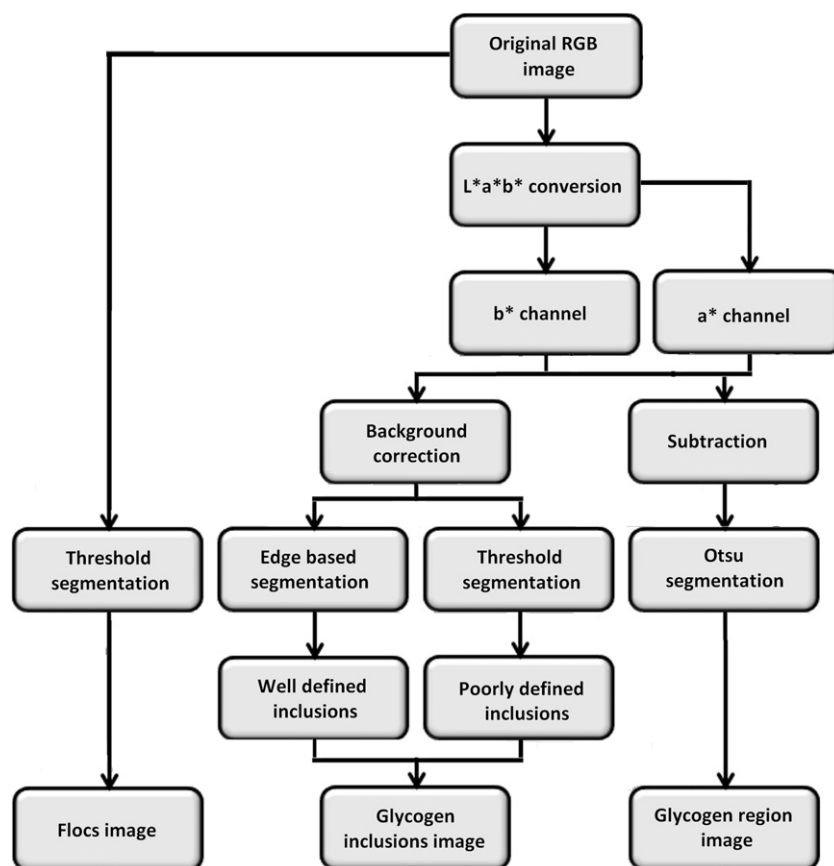


Fig. 1. Schematic representation of the image analysis procedure for intracellular glycogen identification.

Once again the b^* channel was chosen for subsequent treatment. A background correction step, by means of finding and subtracting the minimum of a 10 pixel window around each pixel, was next computed. The identification of the glycogen granules was then performed in a twofold manner: well and poorly defined granules. For the identification of well-defined granules an edge determination and filling procedure was applied, whereas for the poorly defined granules threshold based segmentation was applied. Both well and poorly defined granules were then combined in a single image. The parameters determined included the total glycogen inclusions area, total glycogen granules area and total activated sludge flocs area. For this latter, the original RGB image was segmented to obtain the binary activated sludge flocs image. The percentage of glycogen inclusions and glycogen granules area with respect to the activated sludge flocs area was next computed. Finally, the glycogen inclusions and the glycogen granules binary images were used as mask images in both the original image green RGB channel and the b^* channel, in order to obtain the total and average intensities.

The image processing and analysis methodology regarding PHA inclusions was based on the identification and quantification of PHA inclusion regions. A schematic representation of the image analysis procedure is shown in Fig. 2. The first step of the PHA determination image processing program splits the RGB image into the three composing channels (red, green and blue). Only the red channel was subsequently used for the remaining treatment, given the fact that Nile blue stained PHA inclusions emitted at a red spectra wavelength (625–675 nm). A background correction step, by means of tophat filtering (100 pixel size) was next computed, followed by a wiener filter. The identification of the PHA regions was performed in a twofold manner: well and poorly defined inclusion regions.

For the identification of well-defined inclusion regions a tophat filter (30 pixel size) was applied followed by edge determination and filling, whereas for the poorly defined inclusion regions a threshold based segmentation was applied. Both identified inclusion regions were then combined in a single image that was post-treated in terms of debris elimination by erosion and reconstruction. The final binary image was then used for the determination of the total PHA regions area and subsequently used as a mask image in both the original red channel image and the background corrected red channel image. The parameters total and average inclusion regions intensity were determined for both cases. Finally, the overall image intensity, both in original red channel image and background corrected red channel image, was determined.

Fig. 3 presents the original, fluorescence recognized intracellular storage polymers, and labeled images for both the aniline blue staining (a–c) and Nile blue staining (d–f).

From the collected images, intracellular storage polymers were characterized. Tables 2 and 3 describe the parameters provided by image analysis and subsequently used for the PLS analysis of PHA, PHB and PHV, and glycogen, respectively.

2.7. Multivariate statistical analysis

Partial least squares (PLS) can be seen as a supervised model to assess and monitor any given predicted variable, thus detecting deviations of a certain interest [33]. This method was applied, in the present study, to predict PHA (including PHB and PHV) and glycogen intracellular concentrations using the data from quantitative image analysis as efficiently as possible.

The optimal number of LVs is crucial for the monitoring and prediction outcome and typically cross-validation (leave-one-out)

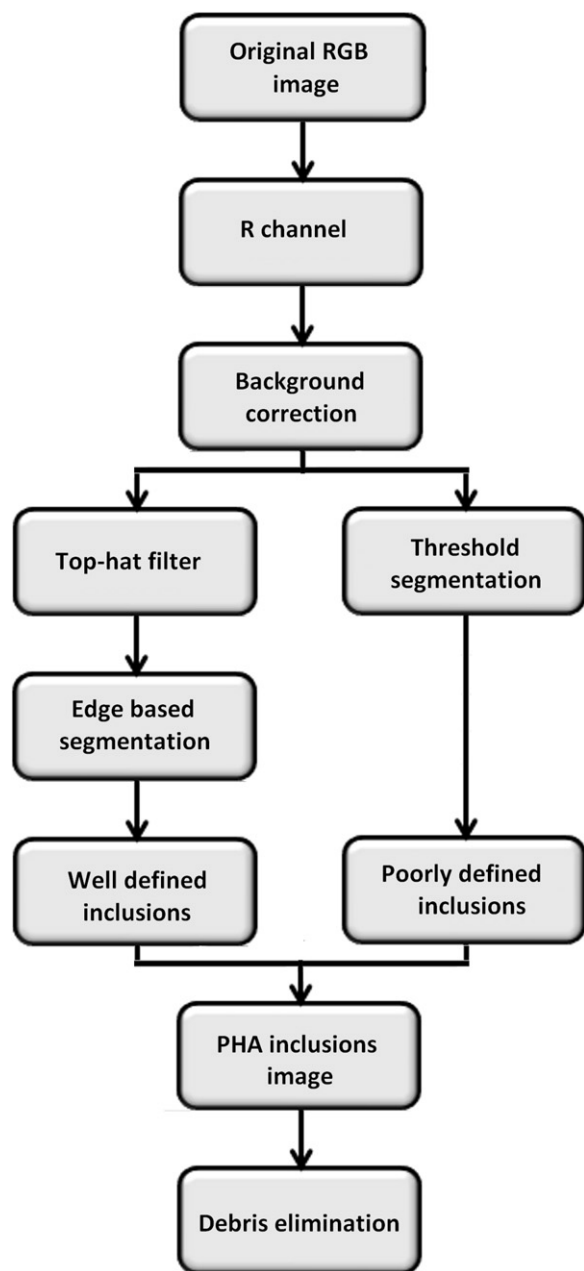


Fig. 2. Schematic representation of the image analysis procedure for intracellular PHA identification.

is used to select the appropriate number, by running sub-model validations where one object (sample) is left out of the data set in each run. In this method, the number of model components (LVs) is usually chosen by the lowest root mean square error of cross-validation, RMSECV [34,35]. A more detailed explanation of PLS algorithms can be found elsewhere [36]. In the current work the maximum number of latent variables that was studied did not surpass half the number of matrix X parameters, for data compression purpose. Furthermore, this number was always considerably smaller than the number of studied samples (Y data points).

However, it was found during the present work that the number of latent variables depicted by this methodology did not allow a satisfactory training model. Furthermore, the use of such methodology may under or overfit the validation data results. To overcome such problems, in this work, the number of latent variables was selected by the difference between the predicted and the measured PHA, PHB, PHV, and glycogen values, expressed by the mean

Table 2
Image analysis variables used for the PLS analysis of PHA, PHB and PHV.

Parameter	Symbol	Description
Predictor (X)	Val.org	Total intensity of the original image
	Val.sbk	Total intensity of the background corrected image
	Area.gran	PHB/PHV granules total area per image
	Total.gran1	PHB/PHV granules total intensity calculated from the original image
	Media.gran1	PHB/PHV granules average intensity calculated from the original image
	Total.gran2	PHB/PHV granules total intensity calculated from the background corrected image
Response (Y)	PHB	Polyhydroxybutyrate concentration
	PHV	Polyhydroxyvalerate concentration
	PHA	Polyhydroxyalkanoates concentration

squared error. Furthermore, and in order to obtain random training and validation datasets, at least 1000 different training and validation datasets from the original dataset were tested. Matrix (X) was always preprocessed using the standard normal variate (SNV) method to remove undesirable variations. With respect to matrix (Y) both the original concentration values and their logarithmic transformation were used to test the model performance. A logarithmic transformation of the Y values was tested due to the fact that the polymers granules are three dimension structures whereas the data obtained from image analysis reflects solely their projection into a two dimensions space.

Matlab™ 7.3 (The Mathworks, Natick, MA) was used for predicting glycogen and PHA (PHB and PHV) intracellular concentrations in an EBPR system by PLS. A first PLS model (designated as overall aerobic and anaerobic) was performed with a total of 108 aerobic and anaerobic observations for PHA (PHB and PHV) prediction and 70 observations for glycogen prediction, divided randomly into a training set (67% of the observations to calibrate the model) and a validation set (33% of the observations to validate the final model). The second study was implemented to anaerobic (AN) and aerobic (AE) data independently, also divided in 67% samples for the training set and 33% for the validation set.

Table 3
Image analysis variables used for the PLS analysis of glycogen.

Parameter	Symbol	Description
Predictor (X)	Area.floc	Flocs total area per image
	Area.lib	Glycogen total area per image
	Area.gran	Glycogen granules total area per image
	Perc.lib	Glycogen area percentage per image
	Perc.gran	Glycogen granules percentage per image
	Total.lib1	Glycogen area total intensity calculated from the original image
	Media.lib1	Glycogen area average intensity calculated from the original image
	Total.lib2	Glycogen area total intensity calculated from the background corrected image
	Media.lib2	Glycogen area average intensity calculated from the background corrected image
	Total.gran1	Glycogen granules total intensity calculated from the original image
	Media.gran1	Glycogen granules average intensity calculated from the original image
	Total.gran2	Glycogen granules total intensity calculated from the background corrected image
Media.gran2	Glycogen granules average intensity calculated from the background corrected image	
Response (Y)	Gly	Glycogen concentration

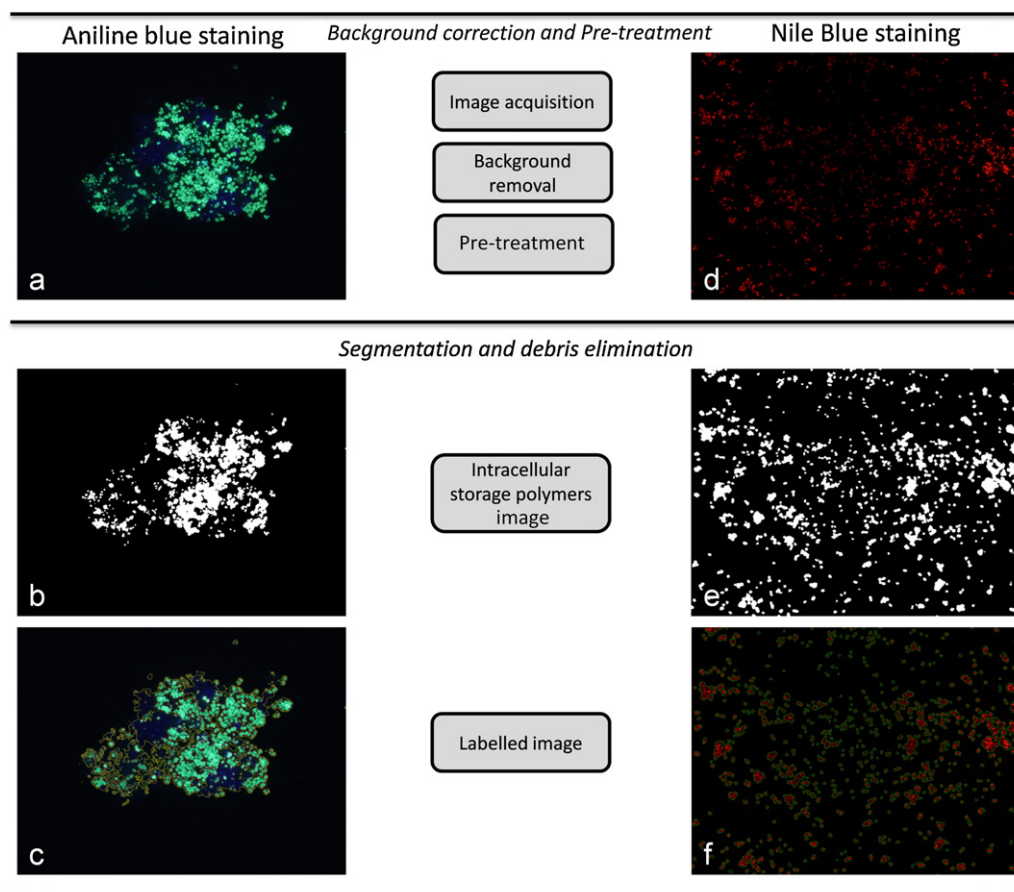


Fig. 3. Images obtained with fluorescence staining procedures at 400× total magnification. (a) Original image from aniline blue staining; (b) binary image with intracellular glycogen identification; (c) labeled image; (d) original image from Nile blue staining; (e) binary image with PHA identification; (f) labeled image. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

3. Results and discussion

The possibility of using quantitative image analysis data to predict intracellular storage polymers through PLS analysis was sought in this work. For this purpose two datasets from the lab-scale EBPR system experiments were used. The first dataset used quantitative image analysis information, provided by fluorescent microscopy from Nile blue staining, to quantify intracellular PHB, PHV, and PHA inclusions (Table 2). Standard analytical methods were used to determine intracellular PHB and PHV concentrations, and it was assumed that PHA concentration was the sum of PHB and PHV. The second dataset combined quantitative image analysis

information, provided by fluorescent microscopy from aniline blue staining, to quantify intracellular glycogen inclusions (Table 3). Standard analytical methods were used to quantify the intracellular glycogen concentration. Both datasets combined results from samples obtained at the end of the anaerobic and aerobic stages.

A first PLS study, using the overall aerobic and anaerobic image analysis data, was performed with the above-mentioned datasets and the PLS analysis results (LVs, regression coefficients (R^2) and linear regression equations) for PHB, PHV, PHA, and glycogen prediction are depicted in Table 4. Both the regression coefficients and regression equations correspond to the linear correlations found by the least squares method. Regarding the predicted and

Table 4
PLS prediction results for PHB, PHV, PHA, and glycogen using the overall aerobic and anaerobic image analysis data.

		Storage polymer							
		PHB		PHV		PHA		Glycogen	
		Y raw data	Pre processing Y	Y raw data	Pre processing Y	Y raw data	Pre processing Y	Y raw data	Pre processing Y
LVs		3	3	3	2	3	3	6	5
R^2 training set		0.33	0.41	0.18	0.26	0.41	0.46	0.81	0.68
Linear regression (training set)	Slope	0.35	0.43	0.18	0.33	0.27	0.47	0.81	0.85
	Intercept	27.9	17.6	13.4	7.8	30.6	20.9	5.8	3.01
R^2 validation set		0.26	0.34	0.40	0.44	0.30	0.33	0.83	0.68
Linear regression (validation set)	Slope	0.34	0.47	0.28	0.62	0.30	0.24	1.2	0.70
	Intercept	24.2	20.2	11.2	7.3	32.1	30.6	-6.4	5.97
R^2 overall (training + validation) set		0.31	0.39	0.25	0.27	0.33	0.40	0.79	0.67
Linear regression (training + validation set)	Slope	0.33	0.44	0.21	0.35	0.28	0.39	0.89	0.81
	Intercept	25.6	18.7	12.6	8.3	31.8	24.7	3.08	3.9

measured intracellular PHB concentrations, low prediction ability was obtained for the crude PHB concentration PLS analysis (Y raw data), presenting a correlation coefficient R^2 of 0.31. The same behavior was obtained for the crude PHV concentration PLS analysis (Y raw data), with even lower prediction ability (R^2 of 0.25). At this point, it is important to note that in EBPR systems containing PAOs and/or GAOs, external acetate is mainly used for PHB production under anaerobic conditions, and to a lesser extent for PHV production [8]. In fact, during the present study, considerably lower PHV concentrations than PHB were obtained (Table 1), and the large majority of samples clustered in a narrow range of values (below 20 mg L^{-1}). Regarding the prediction of intracellular PHA concentrations (determined combining PHB and PHV values), for the crude PHA concentration PLS analysis (Y raw data), only a slightly improvement regarding the PHB prediction was observed, though still presenting low prediction ability (R^2 of 0.33).

These low correlation coefficients, obtained for the PHA, PHB and PHV prediction, for the crude polymer concentrations PLS may suggest that a possible non-linear relationship between these polymer concentrations and image analysis data occurs. As stated earlier, this may be due to the fact that the polymers granules are three dimension structures whereas the data obtained from image analysis reflects solely their projection into a two dimensions space. Thus, a preprocessing method was carried out based on the logarithmic transformation of the polymers concentration values (Preprocessing Y), to improve the model performance. However, the correlation coefficients obtained for PHB, PHV, and PHA prediction abilities demonstrated only a slight improvement (R^2 of 0.39 for PHB, R^2 of 0.27 for PHV, and R^2 of 0.40 for PHA).

Regarding the predicted and measured intracellular glycogen concentrations a satisfactory correlation was obtained (R^2 of 0.79), for the crude glycogen concentration PLS analysis (Y raw data), by aniline blue staining. Although the obtained correlation factor is still somewhat distant from 1, these results can be seen as promising. The preprocessing method used for PHB, PHV, and PHA results was also tested with the glycogen concentrations, however low prediction ability was obtained (R^2 of 0.67) regarding the crude concentrations.

To improve the prediction ability for PHB, PHV, PHA, and glycogen intracellular concentrations, a second study was performed, in which each of the two sampling periods (aerobic and anaerobic) was modeled independently by PLS analysis. The results from these two independent analyses were then combined to obtain the overall PHB, PHV, PHA and glycogen concentrations prediction. The PLS analysis results (LVs, regression coefficients (R^2) and linear regression equations) for PHB, PHV, PHA, and glycogen prediction of this second study are depicted in Table 5. Training, validation, overall training and validation for the aerobic and anaerobic stages, and the combination of the independently obtained aerobic and anaerobic polymers concentrations (depicted as AN+AE) results are presented. Both the regression coefficients and regression equations correspond to the linear correlations found by the least squares method.

Comparing PHB, PHV, and PHA prediction ability from anaerobic and aerobic stages, calculated independently for the crude polymer concentration PLS analysis (Y raw data), the best PLS results were observed for the anaerobic samples (R^2 of 0.55 for PHB, R^2 of 0.48 for PHV, and R^2 of 0.58 for PHA). These results could be explained by the microbial communities' metabolism. In anaerobiosis, intracellular storage polymers (PHA) are stored, thus, the array of concentrations is larger at the end of this stage, opposite to the low concentrations obtained at the end of the aerobic stage, when these polymers have been oxidized. This could be observed in Table 1 concerning the minimum and maximum polymer concentrations obtained during the monitoring period comparing the end of anaerobic and aerobic stages. Moreover, poly-P is accumulated

Table 5
PLS prediction results for PHB, PHV, PHA, and glycogen using independent anaerobic (AN) and aerobic (AE) image analysis data.

	Storage polymer											
	PHB			PHV			PHA			Glycogen		
	Y raw data	Pre processing Y		Y raw data	Pre processing Y		Y raw data	Pre processing Y		Y raw data	Pre processing Y	
AN	AE		AN	AE		AN	AE		AN	AE		
LVs	3	3	2	3	3	3	3	3	2	2	3	3
R^2 training set	0.52	0.27	0.42	0.65	0.39	0.67	0.58	0.13	0.33	0.87	0.86	0.87
Linear regression (training set)	Slope	0.52	0.27	0.37	0.49	0.70	0.58	0.13	0.45	1.27	1.39	0.96
	Intercept	27.33	7.75	37.99	9.17	0.30	33.93	20.61	55.14	4.79	13.08	0.70
R^2 validation set	0.70	0.21	0.73	0.14	0.51	^a	0.63	0.58	0.62	0.91	0.93	0.84
Linear regression (validation set)	Slope	0.80	0.22	0.54	0.52	0.46	0.71	0.22	0.38	0.86	0.86	1.13
	Intercept	17.99	7.88	30.55	10.89	10.64	34.09	18.18	62.97	3.82	4.24	2.69
R^2 overall (training + validation) set	0.55	0.25	0.49	0.55	0.41	0.54	0.58	0.26	0.46	0.86	0.85	0.77
Linear regression (training + validation set)	Slope	0.58	0.25	0.42	0.36	0.61	0.60	0.17	0.43	0.95	0.96	0.84
	Intercept	25.56	7.81	36.59	10.46	9.74	35.24	19.62	57.6	2.00	1.13	1.81
AN + AE R^2 (training + validation set)	0.64	0.68	0.64	0.64	0.53	0.65	0.65	0.69	0.61	0.85	0.85	0.79
Linear regression (AN + AE)	Slope	11.41	0.68	16.82	0.55	0.59	0.69	0.69	0.58	0.96	0.96	0.91
	Intercept				4.16	4.16	17.74	17.74	28.6	1.60	1.60	2.59

^a No correlation coefficient was found.

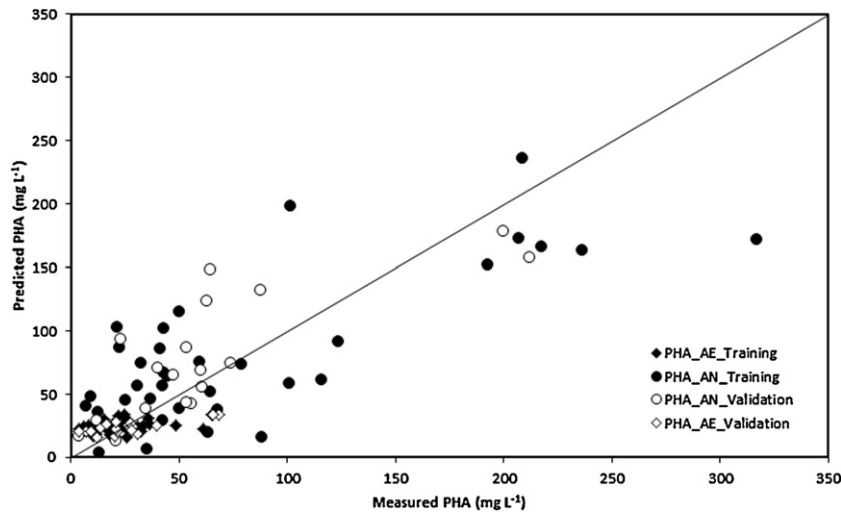


Fig. 4. Measured vs. predicted intracellular PHA concentrations combining anaerobic and aerobic stages, determined independently.

aerobically, and when Nile blue staining is used for PHA inclusions detection, some interferences of poly-P in image analysis procedures could be found depending on the used excitation wavelength.

Also regarding the crude polymer concentration PLS analysis (Y raw data), the comparison between the independent PLS results (AN + AE) (R^2 of 0.64 for PHB, R^2 of 0.42 for PHV, and R^2 of 0.65 for PHA) and the overall datasets results (Table 4) of the first study (R^2 of 0.31 for PHB, R^2 of 0.25 for PHV, and R^2 of 0.33 for PHA), showed a marked improvement. Fig. 4 shows the measured versus predicted values for intracellular PHA concentrations, for (AN + AE) independent PLS results. These results demonstrate thus that a previous identification of the samples stage is the best strategy to predict these intracellular storage polymers.

Again, a preprocessing method was carried out based on the logarithmic transformation of the polymers concentration values (Preprocessing Y), to improve the model performance. And, although, this procedure did not present any improvement regarding the prediction ability of these polymers for the anaerobic stage, it was found to be valuable for the aerobic stage with a significant improvement of the correlation coefficients (R^2 of 0.25–0.55 for PHB, R^2 of 0.16–0.54 for PHV, and R^2 of 0.26–0.62 for PHA). Regarding the preprocessed polymer concentration PLS analysis (Preprocessing Y), the comparison between the independent PLS

results (AN + AE) (R^2 of 0.64 for PHB, R^2 of 0.53 for PHV, and R^2 of 0.61 for PHA) and the overall datasets results (Table 4) of the first study (R^2 of 0.39 for PHB, R^2 of 0.27 for PHV, and R^2 of 0.40 for PHA), also showed a marked improvement. Once again, these results demonstrate the advantages of a previous identification of the samples stage. Nevertheless, the obtained correlation coefficients are still quite far from perfect, although it is known that Nile blue staining has a good capability to stain intracellular PHA (PHB and PHV) storage granules, as demonstrated by the microscopy results obtained for different types of organisms [37,38]. This indicates that the current Nile blue staining and image analysis methodology was not able, so far, to accurately predict these polymers intracellular concentrations. Two main factors may be contributing for these results, worth whiling subsequent study: (1) low biomass fragmentation before microscopic visualization concealing some PHA inclusions; (2) staining and detection of other lipid compounds by the image analysis methodology.

Regarding glycogen inclusions results of the second study (Table 5), for the crude polymer concentration PLS analysis (Y raw data), no significant difference between the anaerobic and aerobic stage prediction abilities was observed, presenting satisfactory correlations (R^2 of 0.86 and 0.85, respectively). During this work, a wide range of glycogen concentrations was obtained for

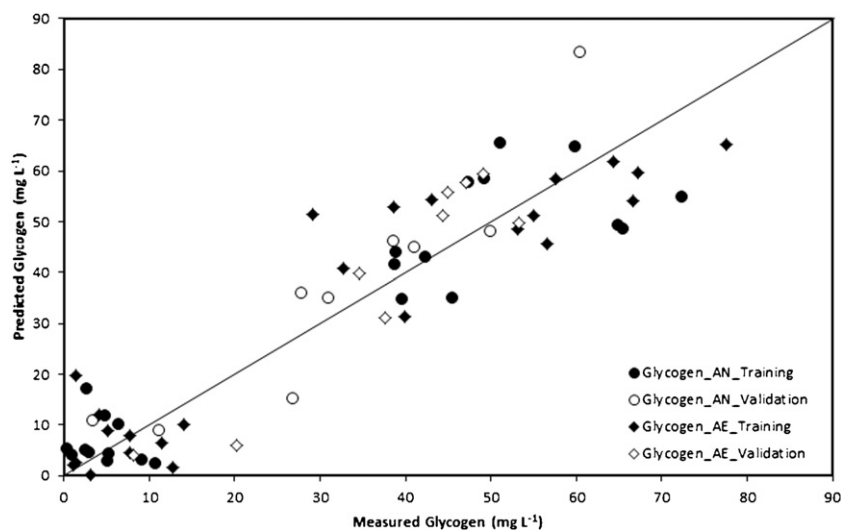


Fig. 5. Measured vs. predicted intracellular glycogen concentrations combining anaerobic and aerobic stages, determined independently.

both the anaerobic and aerobic stages, resulting in a fair approach toward prediction ability in both stages. In fact, taking into account the microbial metabolism, at the end of both anaerobic and aerobic stages, and depending on the state of the sludge, a large range of glycogen concentrations can be obtained for both stages. Nevertheless, comparing the prediction ability of the intracellular glycogen concentrations of the overall dataset of the first study (R^2 of 0.79), to the independent anaerobic and aerobic analysis (R^2 of 0.85), a slight improvement was found. Fig. 5 shows the measured versus predicted values for intracellular glycogen concentrations, for (AN+AE) independent PLS results. These results point to the previous identification of samples stage as the best strategy to predict also glycogen concentrations. Furthermore, the obtained correlation factor strengthen the previous conclusions that the aniline blue staining coupled to image analysis techniques seems to be a promising methodology for intracellular glycogen concentration determination.

With respect to the preprocessing of the glycogen concentration (preprocessing Y), no improvement could be found both for each independent stage as for the combined prediction ability results. Therefore, in the case of glycogen concentrations, it was found no advantage in preprocessing the glycogen concentrations.

4. Conclusions

The ability to predict intracellular glycogen, PHA, PHB, and PHV concentrations measured through off-line analytical measurements, by fluorescent staining and image analysis methodologies was studied in the current work. Concerning intracellular glycogen concentrations, satisfactory prediction abilities were found for both the aerobic and anaerobic stages. Thus, and given the promising results, the use of aniline blue staining coupled to image analysis techniques seems to have the potential to predict glycogen intracellular concentrations to at least some extent. Regarding PHB, PHV, and PHA lower prediction abilities were obtained, and it was concluded that a previous identification of the sample stage (anaerobic or aerobic) in the EBPR system is fundamental. Furthermore, given the lower PHA, PHV, and PHB intracellular concentrations for the aerobic stage, due to the microbial community metabolism, the prediction ability was lower than for the anaerobic stage. Future work will focus on optimizing the PHA, PHB, and PHV staining methodologies and studying larger datasets and different operating conditions, to improve the methodological robustness and prediction ability for the studied intracellular storage polymers.

Acknowledgements

Fundação para a Ciência e a Tecnologia (FCT) in Portugal is gratefully acknowledged by their financial support through the project PTDC/EBB-EBI/103147/2008 and PEst-C/EQB/LA0006/2011. DP Mesquita would like to acknowledge FCT for a post-doctoral grant (SFRH/BPD/82558/2011).

References

- [1] Y. Comeau, D. Lamarre, F. Roberge, M. Perrier, G. Desjardins, C. Hade, R. Mayer, *Water Sci. Technol.* 34 (1996) 169.
- [2] A. Oehmen, R.J. Zeng, Z. Yuan, J. Keller, *Biotechnol. Bioeng.* 91 (2005) 43.
- [3] T. Mino, M.C.M. Van Loosdrecht, J.J. Heijnen, *Water Res.* 32 (1998) 3193.
- [4] R.J. Zeng, M.C.M. van Loosdrecht, Z.G. Yuan, J. Keller, *Biotechnol. Bioeng.* 81 (2003) 92.
- [5] A. Oehmen, A.M. Saunders, M.T. Vives, Z. Yuan, J. Keller, *J. Biotechnol.* 123 (2006) 22.
- [6] T. Mino, M.C.M. van Loosdrecht, J.J. Heijnen, *Water Res.* 32 (1998) 3193.
- [7] J.S. Cech, P. Hartman, J. Wanner, *Water Environ. Res.* 65 (1993) 690.
- [8] A. Oehmen, P.C. Lemos, G. Carvalho, Z.G. Yuan, J. Keller, L.L. Blackall, M.A.M. Reis, *Water Res.* 41 (2007) 2271.
- [9] A. Oehmen, M.T. Vives, H. Lu, Z. Yuan, J. Keller, *Water Res.* 39 (2005) 3727.
- [10] C.M. Lopez-Vazquez, A. Oehmen, C.M. Hooijmans, D. Brdjanovic, H.J. Gijzen, Z. Yuan, M.C.M. Van Loosdrecht, *Water Res.* 43 (2009) 450.
- [11] C.D.M. Filipe, G.T. Daigger, C.P.L. Grady, *Biotechnol. Bioeng.* 76 (2001) 17.
- [12] A.J. Schuler, D. Jenkins, *Water Sci. Technol.* 46 (2002) 171.
- [13] P.C. Lemos, L.S. Serafim, M.A.M. Reis, *J. Biotechnol.* 122 (2006) 226.
- [14] M. Beccari, M. Majone, P. Massaniso, R. Ramadori, *Water Res.* 32 (1998) 3403.
- [15] G.J.F. Smolders, J. van de Meij, M.C.M. van Loosdrecht, J.J. Heijnen, *Biotechnol. Bioeng.* 43 (1994) 461.
- [16] A. Oehmen, R.J. Zeng, A.M. Saunders, L.L. Blackall, J. Keller, Z. Yuan, *Microbiology* 152 (2006) 2767.
- [17] A. Oehmen, C.M. Lopez-Vazquez, G. Carvalho, M.A.M. Reis, M.C.M. van Loosdrecht, *Water Res.* 44 (2010) 4474.
- [18] A.B. Lanham, A.R. Ricardo, M. Coma, J. Fradinho, M. Carvalheir, A. Oehmen, G. Carvalho, M.A.M. Reis, *Biores. Technol.* 118 (2012) 518.
- [19] C. Cenens, K.P. Van Beurden, R. Jenné, J.F. Van Impe, *Water Sci. Technol.* 46 (2002) 381.
- [20] R. Jenné, E.N. Banadda, I.Y. Smets, J. Deurincq, J.F. van Impe, *Microsc. Microanal.* 13 (2007) 36.
- [21] M. da Motta, M.N. Pons, N. Roche, *Water Sci. Technol.* 43 (2001) 91.
- [22] A.L. Amaral, E.C. Ferreira, *Anal. Chim. Acta* 544 (2005) 246.
- [23] D.P. Mesquita, O. Dias, A.L. Amaral, E.C. Ferreira, *Microsc. Microanal.* 16 (2010) 166.
- [24] D.P. Mesquita, A.L. Amaral, E.C. Ferreira, *Chemosphere* 85 (2011) 643.
- [25] D.P. Mesquita, A.L. Amaral, E.C. Ferreira, *Anal. Chim. Acta* 705 (2011) 235.
- [26] D.P. Mesquita, O. Dias, A.M.A. Dias, A.L. Amaral, E.C. Ferreira, *Anal. Chim. Acta* 692 (2009) 94.
- [27] P.L. Bond, R. Erhart, M. Wagner, J. Keller, L.L. Blackall, *Appl. Environ. Microbiol.* 65 (1999) 4077.
- [28] W. Manz, R.I. Amann, W. Ludwig, M. Wagner, F.H. Schleifer, *System. Appl. Micro.* 15 (1992) 593.
- [29] C. Roller, M. Wagner, R. Amann, W. Ludwig, K.H. Schleifer, *Microbiology* 140 (1994) 2849.
- [30] A.G. Ostel, J.G. Holt, *Appl. Environ. Microb.* 44 (1982) 238.
- [31] I.E. Shennawy, D.J. Gee, R.W. Horobin, S.R. Aparicio, *Histochemistry* 81 (1984) 93.
- [32] L.S. Serafim, P.C. Lemos, C. Levantesi, V. Tandoi, H. Santos, M.A.M. Reis, *J. Microbiol. Methods* 51 (1) (2002) 1.
- [33] C. Rosen, *A Chemometric Approach to Process Monitoring and Control with Applications to Wastewater Treatment Operation*. PhD Thesis. Lund Institute of Technology, Sweden, 2001.
- [34] N.D. Lourenço, J.C. Menezes, H.M. Pinheiro, D. Diniz, *Environ. Technol.* 29 (2008) 891.
- [35] M.C. Sarragaça, A. Paulo, M.M. Alves, A.M.A. Dias, J.A. Lopes, E.C. Ferreira, *Anal. Bioanal. Chem.* 395 (2009) 1159.
- [36] J.W. Einax, H.W. Zwanziger, S. Geiss, *Chemometrics in Environmental Analysis*, VCH, Weinheim, Germany, 1997.
- [37] S. Kitamura, Y. Doi, *Biotechnol. Tech.* 8 (5) (1994) 345.
- [38] L. Pierce, M.N. Schroth, *Plant Dis.* 78 (7) (1994) 683.