TÉCNICO LISBOA

Pilot-scale polyhydroxyalkanoate production from food waste assisted by online spectroscopy

Beatriz Vivar de Sousa

Thesis to obtain the Master of Science Degree in

Biological Engineering

Supervisors: Doctor Nídia Dana Mariano Lourenço
Doctor Maria Teresa Ferreira Cesário Smolders

Examination Committee

Chairperson: Prof. Maria Ângela Cabral Garcia Taipa Meneses de Oliveira
Supervisor: Doctor Maria Teresa Ferreira Cesário Smolders
Member of the Committee: Prof. Helena Maria Rodrigues Vasconcelos Pinheiro

November 2019

Declaration

I declare that this document is an original work of my own authorship and that it fulfills all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.

Preface

The work presented in this thesis was performed at the Biochemical Engineering (BIOENG) group, in the Faculty of Sciences and Technology (FCT) of Universidade NOVA de Lisboa (UNL), between March and November, 2019. The thesis was supervised by Dr. Nídia Dana Mariano Lourenço and MSc. Fernando Ramos Silva and co-supervised at Instituto Superior Técnico by Dr. Maria Teresa Ferreira Cesário Smolders. The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements No 773375 and 730349.

Acknowledgements

I would like to thank Dr. Nídia Lourenço, for the scientific supervision and for all the motivational speeches, and Prof. Ascensão Reis for presenting me with the opportunity to develop my work in such a fantastic group. I would also like to thank Dr. Teresa Cesário, for accepting to be my supervisor at IST.

I am extremely grateful to Fernando Silva, for being the most amazing non-official supervisor, and to Bruno Pereira, for being my very best partner in crime all along. I also wish to offer my most sincere gratitude to Kaoutar Aboudi, for showing up at the moment I needed her most and being the most helpful partner ever, and to João Bello, for always making me feel a little bit less alone in this. The warmest thanks to everyone else I had the pleasure to share that pilot-plant with, namely, Cláudia Ralo, Nuno Martins, Cláudia Henriques and Daniela Pequito. Thank you for raising the volume everytime I was feeling down and dancing my stress away. You made everything worthwhile.

To my closest friends, who stood by my side during these crazy five years, I want them to know they were the absolute best part of it. To Catarina Soares and Renato Castelão for being the two best partners and friends I could have wished for; to Guilherme Fernandes, Rafaela Pereira, Sofia Navalho, Mafalda Santos and Pedro Silva for being the amazing, unique and incredibly fun people they are, and to Marta Costa and Diogo Abreu, for teaching me a whole new concept of family.

I also wish to extend my gratitude to all of my non-engineering friends, for bringing some balance into my life and helping keep me alive and sane through this period. To Gonçalo Pinto, Tiago Franco, Joana Paulino, Luísa Ventura, Catarina Silva, Laura Ehm, Laura Silvério, Henrique Carvalho and Simão Virgílio.

My most special thanks to my family, as they are the main reason I've come this far. To my dad and my mom, for always pushing me to do nothing but my best and for presenting me with all the amazing opportunities I've had throughout the years. To my sister, for being my best friend in the whole world and for taking care of me when I was drowning in stress. To the rest of my family for always being so incredibly supportive and taking so much pride in me.

Lastly, and above anyone else, I want to thank my boyfriend, Marcelo Paulino, who was absolutely fundamental during these months. His endless patience, encouragement and kindness made everything incomparably easier and better. I will be forever grateful for having him by my side.

Resumo

Perante a atual crise de acumulação de plásticos, polímeros biobaseados e biodegradáveis podem ser preponderantes na construção de uma indústria de plástico mais sustentável. Entre estes polímeros "verdes", os polihidroxialcanoatos (PHA) são poliésteres alifáticos sintetizados por bactérias e armazenados como grânulos intracelulares. No presente estudo, poli-(3-hidroxibutirato-co-3-hidroxivalerato) foi produzido por uma cultura microbiana mista alimentada com resíduo de polpa de fruta fermentado. O reator de acumulação foi monitorizado *at-line* por espetroscopia Raman e de infravermelho-próximo (NIR) para desenvolver modelos da concentração de sólidos suspensos totais (SST), concentração de PHA e conteúdo de PHA na biomassa. Uma análise de componentes principais (PCA) aos espetros de Raman revelou uma mudança gradual na forma do sinal com o tempo, que se acredita resultar de alterações na cultura. Os espetros de NIR foram utilizados no desenvolvimento de modelos quantitativos através de regressão de mínimos quadrados parciais (PLS). Os modelos, submetidos a validação cruzada e externa, mostraram-se adequados para previsão dos parâmetros estudados. A concentração de SST foi prevista com raíz do erro quadrático médio de previsão (REQMP) de 1.0g/L e coeficiente de determinação (R^2) de 0.86, enquanto que a concentração de PHA apresentou REQMP de 0.69g/L e R^2 de 0.89 e o conteúdo de PHA revelou REQMP de 14.6% e R^2 de 0.86. Finalmente, os resultados dos modelos para concentração de SST e PHA foram combinados para dar uma previsão indireta do conteúdo de PHA. Este modelo mostrou-se excepcional para amostras com elevada densidade celular, com REQMP de 4.52% e R^2 de 0.96.

Palavras-chave

Polihidroxialcanoatos; Culturas Microbianas Mistas; Resíduos de Polpa de Fruta; Monitorização de Reatores Biológicos; Espetroscopia Raman; Espetroscopia de Infravermelho Próximo; Análise Multivariada.

Abstract

Given today's plastic disposal global crisis, bio-based and biodegradable polymers may have an important role in leading the way towards a more sustainable plastics industry. Among these "green" polymers, polyhydroxyalkanoates (PHA) are aliphatic polyesters synthesized by bacteria and stored as intracellular granules. In this study, poly(3-hydroxybutyrate-co-3-hydroxyvalerate), P(3HB-co-3HV), was produced by a mixed microbial culture fed with fruit pulp waste fermentate. The accumulation reactor was monitored at-line by Raman and NIR spectroscopy to develop models for predicting total suspended solids (TSS) concentration, PHA concentration and intracellular PHA content. A principal component analysis (PCA) model performed on the Raman spectral data revealed a gradual change in the shape of the signal over time, which was believed to result from shifts in the bacterial community. NIR spectra were used for the development of quantitative calibration models through partial least squares (PLS) regression. The models were subjected to cross-validation and external test set validation and proved to be suitable for predicting all three parameters under study. TSS concentration was predicted with a root mean square error of prediction (RMSEP) of 1.0g/L and a coefficient of determination (R^2) of 0.86, while PHA concentration showed a RMSEP of 0.69g/L and a R^2 of 0.89 and PHA content had an RMSEP of 14.6% and a R^2 of 0.86. Finally, the results of the TSS and PHA concentration models were combined to give an indirect PHA content prediction. This model performed exceptionally in predicting high cell density samples, with a RMSEP of 4.52% and a R^2 of 0.96.

Keywords

Polyhydroxyalkanoates; Mixed Microbial Culture; Fruit Pulpe Waste; Bioreactor Monitoring; Raman Spectroscopy; Near-Infrared Spectroscopy; Multivariate Analysis.

Contents

List of Tables

List of Figures

List of Acronyms

AD	Anaerobic Digestion
ADF	Aerobic Dynamic Feeding
C/N	Carbon/Nitrogen ratio
COD	Chemical Oxygen Demand
DO	Dissolved Oxygen
EBPR	Enhanced Biological Phosphorus Removal
EU	European Union
FBR	Fed-Batch Reactor
F/F	Feast/Famine ratio
FISH	Fluorescence *In Situ* Hybridization
GAO	Glycogen-Accumulating Organism
GC	Gas Chromatography
GLOPACK	Granting Society with LOw Environmental Impact Innovative PACKaging
HA	Hydroxyalkanoate
HB	Hydroxybutyrate
HD	Heptadecanoate
HHD	Hydroxyhexadecanoate
HHx	Hydroxyhexanoate
HPLC	High-Performance Liquid Chromatography
HRT	Hydraulic Retention Time
HV	Hydroxyvalerate
LCL	Long-chain-length
MCL	Medium-chain-length
MIR	Mid-Infrared
MMC	Mixed Microbial Culture
NGO	Non-Governmental Organization

NIR	Near-Infrared
OLR	Organic Loading Rate
PA	Polyamide
PAO	Polyphosphate-Accumulating Organism
PBAT	Polybutyrate Adipate Terephthalate
PBS	Polybutylene Succinate
PC	Principal Component
PCA	Principal Component Analysis
PE	Polyethylene
PET	Polyethylene Terephthalate
PFA	Paraformaldehyde
PHA	Polyhydroxyalkanoate
PLA	Polylactic Acid
PLS	Partial Least Squares
PP	Polypropylene
PS	Polystyrene
P(3HB)	Poly(3-Hydroxybutyrate)
P(3HV)	Poly(3-Hydroxyvalerate)
P(3HB-co-3HV)	Poly(3-Hydroxybutyrate-co-3-Hydroxyvalerate)
P(3HB-co-3HHx)	Poly(3-Hydroxybutyrate-co-3-Hydroxyhexanoate)
SBR	Selection Batch Reactor
SCL	Short-chain-length
SDS	Sodium Dodecyl Sulphate
SG	Savitzky-Golay
SNV	Standard Normal Variate
SRT	Solids Retention Time
TSS	Total Suspended Solids
UASB	Upflow Anaerobic Sludge Blanket
UN	United Nations
UV	Ultraviolet
VFA	Volatile Fatty Acid
WWTP	Wastewater Treatment Plant

1 Introduction

1.1 Context of the Thesis

We are currently facing a society that is strongly dependent on plastics throughout many sectors: from packaging to agriculture, electronics, construction industry, transportation or health care. Due to their inexpensive manufacturing, high resistance and versatility, plastics have become an integral and important part of the global economy [1]. They began to be commercialized around 1950's and their global production reached an astounishing number of 348 milllion tonnes in 2017, making them a potentially unparalleled success in the history of materials industry [2, 3].

However, at a time in which a global environmental conscience is arising, the issues associated with the origin and fate of conventional plastics cannot possibly be ignored. On the one hand, over 99% of all commercially available plastics are produced from fossil fuels - if current trends continue, by 2050 the plastic industry could account for 20% of the world's total oil consumption [4]. On the other hand, their non-biodegradability, together with inefficient collection and recycling systems, is leading to an alarming accumulation of mismanaged plastic waste. This has been aggravated by the fact that, over the last decades, production has shifted from durable to single-use plastic [1, 3].

In Europe, one of the world regions with the most effective waste management systems, only 31% of collected plastic was recycled in 2016, with 42% used for energy recovery through incineration and the remaining 27% ending up in landfills [2]. This means that the potential for recycling plastic waste remains largely unexploited.The scenario is significantly more negative if a global and cummulative analysis is carried out: 5.8 billion tonnes of plastic waste (polymers, synthetic fibers and additives) were produced between 1950 and 2015, of which 84% was discarded, 14% was incinerated and only 2% was recycled and is still in use [5]. Furthermore, it is estimated that between 5 and 13 million tonnes of plastic end up in the ocean every year, endangering marine environments at levels yet to be fully understood [6].

Packaging, in particular, accounts for about 40% of global annual plastic production and is almost exclusively single-use [4, 5]. It is estimated that only 14% of plastic packaging is collected for recycling and that after sorting and reprocessing only 5% of material value is retained. Furthermore, plastics that do get recycled are mostly recycled into lower-value applications that are no longer recyclable[4]. Plastic packaging is also among the items most commonly found on beaches, since around 32% is believed to escape collection systems [4, 7].

Given the current scenario, in January 2018, the European Union (EU) presented its strategy for a more sustainable plastics industry, which relies on a circular economy concept and is to be achieved by 2030 [7]. Besides promoting and optimizing plastic recycling, the report discusses opportunities and risks of bio-based and biodegradable plastics (see Section 1.2.1). Moreover, 'The New Plastics Economy Global Commitment' was signed in Bali, in October 2018, and focuses on innovating to ensure 100%

of plastic packaging can be safely reused, recycled, or composted by 2025 [8, 9]. The initiative is led by the Ellen MacArthur Foundation, in collaboration with UN Environment, and is supported by many of the world's largest packaging producers, brands, retailers and recyclers, as well as governments and non-governmental organizations (NGOs).

Therefore, it is becoming widely accepted that a possible approach to the current plastic waste crisis might be the further development and production of biorelated plastic materials. These alternative polymers, may, in fact, present themselves as answers to the environmental problems raised by traditional plastics, both reducing nonrenewable energy use and, consequently, greenhouse gas emissons, as well as presenting an innovative solution for the end of life issues, as will be discussed in subsequent sections [10].

A number of projects has been funded by the EU over the last few years on the field of bio-based and biodegradable polymers [11, 12, 13, 14, 15]. The GLOPACK (Granting society with LOw environmental impact innovative PACKaging) project is investigating food packaging with little environmental footprint and the ability to extend the shelf life of food products. It has over a dozen partners all over Europe, from research groups and universities to private companies. It is within this project that the present thesis is inserted, focusing on biodegradable packaging materials made from agro-food residues [15].

1.2 Biorelated Polymers

1.2.1 Classification and Properties

Biorelated polymers are generally associated with two different concepts: 'bio-based' and 'biodegradable'. The term 'bio-based' implies the material contains organic carbon of renewable origin, i.e. it is total or partially derived from biomass, whereas 'biodegradable' means it can be converted into natural substances such as water, carbon dioxide/methane and nitrate/ammonium by the action of naturally-occurring microorganisms [16, 17]. Contrary to common belief, biodegradation is unrelated to the origin of a material, but is rather a consequence of its chemical structure. This means that bio-based plastics may not be biodegradable, just like plastics based on fossil fuels may be susceptible to biodegradation [18].

Bio-based polyethylene (PE), polyethylene terephthalate (PET) and polyamide (PA), for instance, are similar to their their oil-based counterparts in terms of molecular structure but are derived from renewable feedstock, i.e., they are bio-based. This means they exhibit the same properties and are suitable for the same applications, but produce a smaller carbon footprint at their source, even though they do not show any biodegradability capacity [16]. On the other hand, polybutylene succinate (PBS) and polybutylene adipate-co-terephthalate (PBAT) are examples of petrochemical-based polymers that given their chemical structure are susceptible to biodegradation at their end of life [19].

The polymers that have, however, received greater attention and show more promise as truly sustainable alternatives to conventional plastics are those that are both bio-based and biodegradable, since they tackle the two major environmental issues associated with the latter. Three main polymers are included in this group: polysaccharides (e.g. starch), polylactic acid (PLA) and polyhydroxyalkanoates (PHA).

Apart from biodegradability, if these biorelated polymers are to be suitable for replacing traditional plastic materials, they must exhibit specific features, depending on the particular application and the

oil-based plastic they aim to replace. These include mechanical (e.g. elasticity, strenght), thermal (e.g. melting point, glass transition temperature) and optical (e.g. transparency) characteristics, as well as gas and water barrier properties. Furthermore, it is desirable that these polymers are thermoplastically processable through conventional methods, such as extrusion, molding and foaming, in order for their introduction in the market to be as smooth and cost-efficient as possible [19].

Starch

Starch is the most abundant polysaccharide reserve in plants, therefore representing a renewable, biodegradable and readily available basis for polymer development. In fact, starch can be extracted from cereals, tubers, grains and even nuts, with its main commercial sources coming down to corn, potato and tapioca [20].

Even though starch polymers from different sources will exhibit different properties, starch is hardly used in its native state, as it is generally brittle, absorbs moisture and is difficult to process. Its properties can be improved through the addition of plasticisers (e.g. sorbitol and glycerine) or through blending with a range of other polymers (e.g. PE, PHA, PLA), even though sometimes compromising its biodegradability [16, 21, 22]. The resulting starch-based thermoplastic materials have been successfully applied on industrial level for foaming, film blowing, injection molding, blow molding and extrusion applications and used in several industries, such as food, medical and pharmaceutical [20].

Polylactic acid (PLA)

PLA is a biodegradable aliphatic polyester which can be produced from bio-based lactic acid monomers resulting from the fermentation of renewable resources such as corn, wheat, potato and sugarcane [23]. Polymerization is achieved through classical chemical synthesis, either by direct polycondensation or ring-opening polymerization [24].

PLA exhibits mechanical, optical and barrier properties comparable to those of oil-based commodity polymers, such as polypropylene (PP), PET, and polystyrene (PS). Furthermore, processing of PLA can now be carried out through most well-established polymer-manufacturing techniques, from injection molding to film extrusion. Consequently, PLA has potential applications in a wide range of fields, from biomedical devices and implants to packaging and agricultural industries [25].

Although it may take several years for PLA to completely disappear in a natural environment, it is recyclable and quickly biodegradable under industrial composting conditions. It is mostly suitable in cases where plastics become highly contaminated and are difficult to recover through traditional recycling systems (e.g. food packaging containers, agricultural films). However, suitable infrastructures for sorting, recycling, and/or composting PLA products at their end of life must still be developed to enable this biopolymer to perform to its true potential [23, 25].

Polyhydroxyalkanoates (PHA)

Polyhydroxyalkanoates (PHAs) are a family of aliphatic polyesters produced via bacterial metabolic transformation of sugars or fatty acids, generally plant-based. In native PHA-producing organisms, they function as intracellular reserves for carbon and energy. They are also easily biodegraded by a range of bacteria that secrete specific extracellular PHA depolymerases [16, 19]. The most relevant

polymers belonging to this family are poly(3-hydroxybutyrate) (P(3HB)), poly(3-hydroxybutyrate-co-3-valerate) (P(3HB-co-3HV)) and poly(3-hydroxybutyrate-co-3-hydroxyhexanoate) (P(3HB-co-3HHx)), all of which are already commercially available [26, 27, 28, 19].

PHA's characteristics are dependent on the type of carbon source and the microorganisms involved, which determine the composition of the polymer [22]. In fact, by manipulation of the molecular structure, properties ranging from hard, brittle plastics to soft elastomers can be obtained, using conventional processing techniques [19]. This means they show potential to be used as alternatives for several traditional polymers and applied to a range of different industries [22].

Polyhydroxyalkanoates will be the main focus of this thesis, meaning that more detailed analysis will be presented in subsequent sections, namely regarding their properties, bacterial metabolism, production process and applications.

1.2.2 Market

In 2017, an astounishing amount of 348 million tonnes of plastic were produced globally [2]. On the same year, the global production of biorelated polymers (bio-based and/or biodegradable) summed up to a mere total of 2.06 million tonnes, and 2.11 million on the following year. This means that "bioplastics" currently account for less than 1% of annual plastic production, worldwide [29].

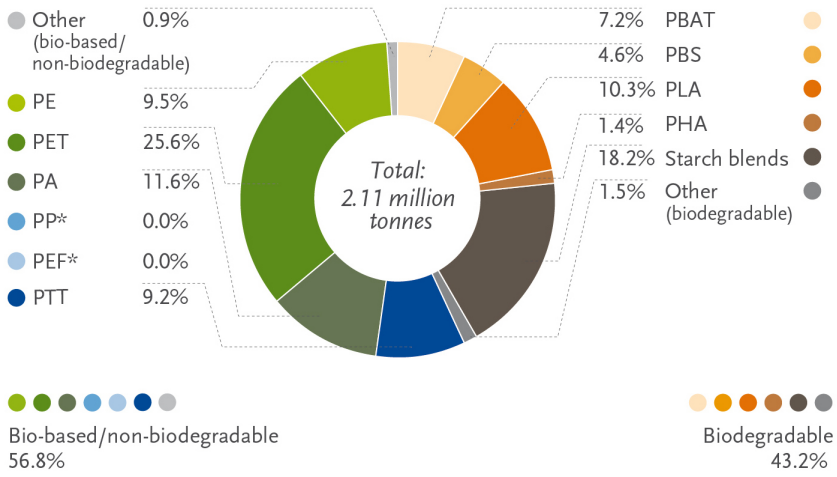


Figure 1.1: Global production capacity of biorelated polymers (by material type) in 2018. PE (polyehtylene), PET (polyethylene terephthalate), PA (polyamide), PP (polypropylene), PEF (polyethylene furanoate), PTT(polytrimethylene terephthalate), PBAT (polybutylene adipate-co-terephthalate), PBS (polybutylene succinate), PLA (polylactic acid), PHA (polyhydroxyalkanoate). * currently under development and predicted to be available at commercial scale in 2023. Source: European Bioplastics, nova-Institute (2018) [29].

Bio-PET (26.6%) and starch blends (18.2%) represent the biggest fraction of the 2.11 million tonnes of bio-based and biodegradable polymers, respectively, produced in 2018. Although PHAs are still a minor share of the global production capacity (1.4%), they are the polymer family that is expected to grow the most over the next five years. In fact, PHAs have been in development for a while and now seem to be ready to enter the market at commercial scale, with production capacities expected to triplicate by 2023 compared to 2018, whereas PLA is estimated to almost double its numbers. Furthermore, biodegradable polymers are expected to grow faster than bio-based. The above mentioned information can be visualized in Figures 1.1 and 1.2 [30].

As a whole, the biopolymer industry global capacity is predicted to reach 2.62 million tonnes in the next five years [29], which is still under 1% of the global plastic production. However, demand for sustainable products is on the rise and improved materials, processes and applications are emerging every year, promising a greener future for the plastic industry.

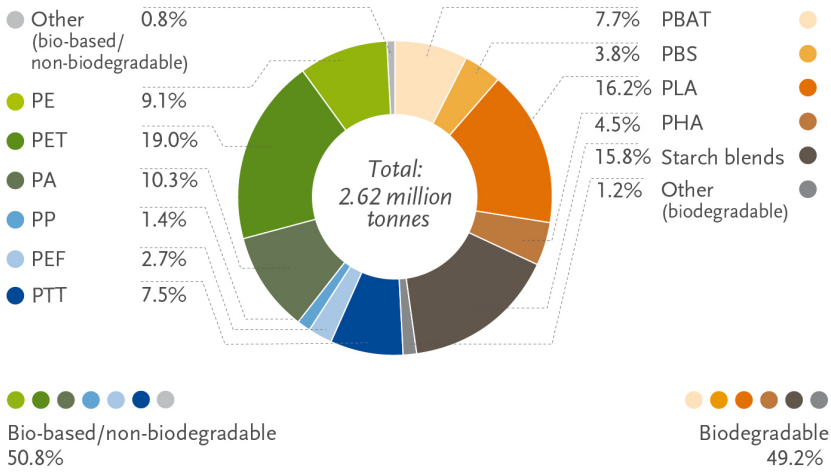


Figure 1.2: Predicted global production capacity of biorelated polymers (by material type) in 2023. PE (polyehtylene), PET (polyethylene terephthalate), PA (polyamide), PP (polypropylene), PEF (polyethylene furanoate), PTT(polytrimethylene terephthalate), PBAT (polybutylene adipate-co-terephthalate), PBS (polybutylene succinate), PLA (polylactic acid), PHA (polyhydroxyalkanoate). Source: European Bioplastics, nova-Institute (2018) [29].

Regarding applications, there is currently a bio-based/biodegradable alternative for almost every conventional plastic material and corresponding application [29]. Therefore, these biopolymers are now used in an increasing number of markets, with packaging taking the lead for both bio-based and biodegradable plastics, with over 50% of the total market in 2018. One can also clearly see that depending on the application and properties required, bio-based or biodegradable plastics are preferred. For instance, biodegradable polymers are mostly used for flexible packaging (e.g. food packaging) and agriculture and horticulture (e.g. agricultural films), where biodegradability is preponderant, as mentioned before. On the other hand, bio-based polymers are used for rigid packaging and more durable products for the textile, building and automotive industries, which require mechanical properties that are still somewhat difficult to achieve with polymers such as PLA or PHA. Figure 1.3 illustrates the previous statements, providing data about the different market sectors on which biorelated polymers are succeeding.

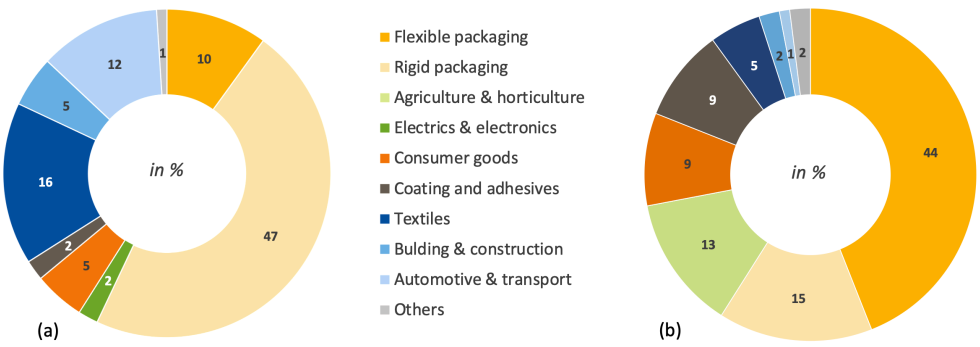


Figure 1.3: Bio-based (a) and biodegradable (b) plastics by market segment in 2018. Source: European Bioplastics, nova-Institute (2018) [29].

1.3 Polyhydroxyalkanoates Fundamentals

1.3.1 Structure and Properties

It was back in the 1920s when French microbiologist Maurice Lemoigne first came across the polyester poly(3-hydroxybutyrate), P(3HB), in the form of intracellular granules in *Bacillus megaterium* [31]. Nowadays, 3HB is still by far the most commonly observed monomer, although over 150 hydroxyalkanoate (HA) monomers and over 300 PHA-producing bacterial species have now been identified [10].This wide range of possible monomeric compositions, microstructure and molecular weight distribution grants PHAs the flexibility in material properties that PLA and starch-based plastics lack [31].

Apart from the range of possible polymers and properties, one of the major advantages of PHA is the fact that it is industrially compostable, home compostable, soil and marine biodegradable, anaerobically digestible and biodegrades in wetlands, septic systems and municipal solid waste systems, unlike PLA which only biodegrades under very specific conditions like industrial composting [19].

Structurally, PHAs can be classified based on the number of carbon atoms in their monomeric units. Short-chain lenght PHA monomers (SCL), like 3-hydroxybutyrate (3HB) and 3-hydroxyvalerate (3HV), contain between 3 and 5 carbon atoms, whereas medium-chain lenght (MCL), like 3-hydroxyhexanoate (3HHx), contain up to 14 carbon atoms [32]. Long-chain lenght (LCL) with over 14 carbon atoms, like 3-hydroxyhexadecanoate (3HHD), are less common but have been identified in wild strains [33]. Regardless of the monomer type, PHA polymerization is achieved by the formation of an ester bond between the carboxyl group of one hydroxy fatty acid and the hydroxyl group of the adjacent monomer [16].

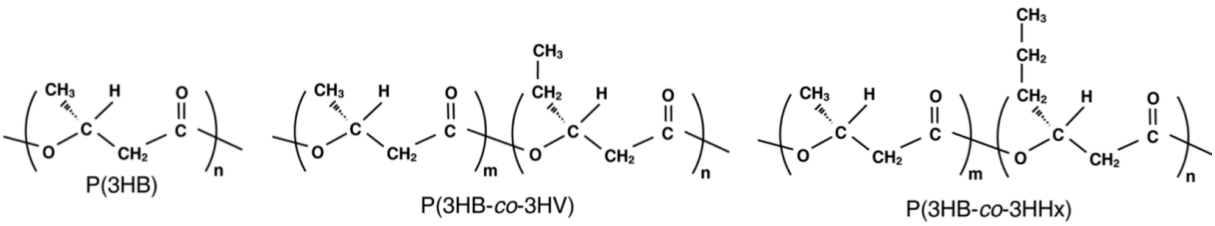


Figure 1.4: Chemical structure of the most well-known polyhydroxyalkanoates: homopolymer poly(3-hydroxybutyrate) (P(3HB)), copolymer poly(3-hydroxybutyrate-co-3-hydroxyvalerate) (P(3HB-co-3HV)) and copolymer poly(3-hydroxybutyrate-co-3-hydroxyhexanoate) (P(3HB-co-3HHx))

Regarding their monomeric composition, PHAs can be classified as homopolymers or copolymers. Homopolymers composed solely of SCL monomer units generally have thermoplastic properties, while those comprehending MCL units generally tend to have elastomeric properties [32]. When copolymers are introduced, however, the possibilities to develop novel polymers with different physical properties become almost endless. figure 1.4 illustrates the chemical structure of some of the most relevant PHA polymers.

P(3HB), for example, exhibits good thermoplastic properties, with a melting temperature of almost 180°C, a glass transition temperature around 4°C and mechanical properties close to those of PP [34]. However, it has a highly crystalline structure (55-85%) that makes it stiff and brittle, with an elongation to break of 5%, in comparison to 400% achieved by conventional PP [10, 31]. In order to overcome this limitation, which makes P(3HB) unsuitable for many applications, different monomers can be incorporated to form copolymers with improved mechanical properties, as they reduce polymer crystallinity by disturbing

the crystal lattice. In P(3HB-co-3HV), for instance, increasing the HV content in the copolymers reduces the melting temperature and crystallinity, while providing toughness and flexibility to the polymer, which facilitates further chemical or mechanical processing [35, 36]. Furthermore, the introduction of MCL monomers, like in P(3HB-co-3HHx), causes even further crystal disruption and yields highly amorphous elastomers for high value-added application, such as surgical sutures, implants, biodegradable matrices for drug delivery, etc.[10, 37].

The monomeric composition of PHA polymers can be influenced by different factors, such as the organism producing the PHA polymer, the carbon source on which cells are grown and how that carbon source is metabolized in the cells, all of which will be discussed further on.

1.3.2 Organisms for PHA Production

Pure Cultures

PHA-producing microorganisms can be found in nearly every conceivable habitat, from hypersaline ecosystems and marine niches to rhizospheric soil and man-made environments, such as waste water treatment plants (WWTP) [38]. Some of the most well-studied and industrially exploited species include *Pseudomonas putida*, *Cupravidus necator*, and *Alcaligenes latus* [37, 39, 40, 41].

Since PHA functions as a carbon storage and energy reserve, its biosynthesis by bacteria usually takes place when the cell is growing at a slower rate than the one at which it takes up its source of carbon. The growth limitation can either result from external lack of nutrients (e.g. nitrogen or phosphorus) or from internal inhibition of an anabolic enzyme [10]. However, there are some bacterial strains, such as *Alcaligenes latus*, which do not require any growth limitation in order to store PHA [31].

Currently, industrial PHA production is conducted by bacterial pure culture fermentation using pure substrates (i.e. glucose), sterile conditions and a fed-batch regime. The general production process can be summarized in two main steps: firstly, the culture is supplied with growth media; secondly, once an appropriate cellular density is reached, growth limiting conditions are imposed and PHA storage is induced [10, 31]. In one of the most successful cases ever reported, this strategy was applied to *A. latus* with nitrogen limitation and allowed for cell densities above 100g/L, P(3HB) contents up to 88% and productivities near 5g/(L.h) [42].

Despite the above mentioned successes, production costs remain too high for this process to replace that of petrochemical commodity polymers. These costs are mainly driven by the market price of pure substrates, the downstream process and the high energy expenditure used in sterilization, aeration and agitation, all of which are also responsible for a significant environmental impact [43].

Over the past decades, there has been an increasing effort towards developing low-cost alternative processes. Strategies have focused, namely, on genetic engineering tools to design improved PHA-producing strains and on the use of agro-industrial wastes and byproducts as substrates for PHA production [38]. Even though these low-cost carbon sources have been tried on pure cultures and good productivities and PHA contents have been demonstrated [44, 45, 46], their fullest potential is only achieved when combined with mixed microbial cultures (MMC), which will be further discussed [10].

Genetically Engineered Strains

Engineering of PHA-producing candidates has been used as a tool to optimize their biosynthetic performance and reduce operation costs in several ways. Firstly, genetic information associated to PHA accumulation can be transferred from wild-type producers to well-known recombinant organisms, such as *Escherichia coli*, whose physiology, biochemistry and genetics have been thoroughly studied. In fact, since some strains of *E. coli* utilize lactose, which many PHA-producing organisms like *C. necator* are not able to metabolize, recombinant *E. coli* strains have been designed with the PHA-producing genes from *C.necator* to produce PHA using dairy whey as a carbon source, with very high PHA contents (87%) [47]. While whey is the primary food waste substrate investigated for recombinant *E. coli*, other nutrient sources and recombinant strains have been investigated [48, 49]. On the other hand, native PHA-producing microrganisms can also be engineered to metabolize additional substrates [50]. Both strategies attempt to tackle the high costs associated with pure substrates [51].

Furthermore, metabolic engineering has allowed for the redirection of substrates towards PHA production in *Pseudomonas putida*, by silencing alternative biosynthetic pathways, therefore increasing the yields of carbon to PHA conversion [52]. The same bacterial strain has also been used as a platform to produce PHA with specific (3HB) and (3HHx) content and composition, providing a solution to batch variability in wild-type cultures [51, 53].

Other optimizations have been performed, from reducing downstream costs by facilitating recovery of PHA granules after cell lysis to accelerating cell growth, achieving very high PHA contents and engineering extremophiles able to grow under open unsterile conditions with high substrate to PHA conversion efficiency [54, 55]. Future advancements in synthetic biology, genetic and metabolic engineering will certainly open an array of opportunities for a more efficient, low-cost and environmentally responsible PHA production [56, 51]. It is worth mentioning, however, that dealing with genetically engineered strains raises some concerns regarding their disposal into the environment, since these are not naturally-occurring microorganisms and might, therefore, affect native ecosystems [57].

Mixed Microbial Cultures (MMC)

Another alternative approach to the reduction of PHA production costs is to replace wild-type or engineered strains with mixed microbial cultures (MMCs). These consortia are microbial populations operating in open (non-sterile) biological systems to which a selective pressure that favors organisms with high PHA storage capacity is applied. In other words, the feeding and cultivation conditions imposed on the bioreactor are such that a bacterial population rich in PHA-producing microorganisms is eventually obtained. Thus, the ecosystem, rather than a specific microorganism, is engineered [10, 37, 38, 58].

The fact that the process takes place under non-sterile conditions and the potential of MMC to adapt to a variety of different inexpensive waste streams as carbon sources are the two main factors behind the cost reduction of producing PHA with MMCs instead of pure cultures [37]. This production strategy is the starting point of the present thesis and will therefore be discussed in greater detail over the next chapter, together with the different agro-industrial/domestic feedstocks used, the most common operating conditions for ecological selection and some notable achievements and applications so far.

1.3.3 Metabolism

The most common pathway for PHA biosynthesis in pure cultures (e.g. *C. necator* and *A. latus*) begins with carbohydrate catabolic degradation through glycolysis. Under unlimited growth conditions, the resulting pyruvate is converted to acetyl-CoA, which enters the tricarboxylic acid cycle. However, when some external or internal limitation to growth is present, acetyl-CoA can serve as a precursor for the synthesis of hydroxyalkanoate monomers. In the latter case, three enzymes are involved: 3-ketothiolase catalyzes the condensation of two units of acetyl-CoA into acetoacetyl-CoA; acetoacetyl-CoA reductase reduces it to (R)-3-hydroxybutyryl-CoA, which is then polymerized as (3HB) into P(3HB), by PHA synthase [10, 59].

However, the feeding of CH-rich feedstocks in a MMC process, such as starch or cellulose hydrolisates, generally results in a culture dominated by glycogen-accumulating organisms (GAOs). Therefore, volatile fatty acids (VFAs) such as acetic, propionic, butyric and valeric, are generally preferred. Due to the fact that not much is known about the metabolism for PHA production by MMCs, it is assumed that it is similar to that reported for pure cultures using the same carbon substrate [59].

Regarding the direct use of VFAs, acetate is converted to acetyl-CoA, then to (R)-3-hydroxybutyryl-CoA, and finally to PHB, through the same chemical reactions and using the same enzymes as pure cultures. Similarly, propionate is converted to propionyl-CoA, which can then be combined with an acetyl-CoA molecule to form (R)-3-hydroxyvaleryl-CoA, and finally PHV. This second pathway is analogous to the one for PHB biosynthesis, with the only difference being that propionyl-CoA replaces one of the acetyl-CoA molecules [60]. It is worth noticing that propionyl-CoA can also lead to acetyl-CoA via decarboxylation, meaning that the presence of acetate is not determinant for PHB synthesis [60]. Butyrate and valerate, which have a higher number of carbon atoms, can be directly converted to (R)-3-hydroxybutyryl-CoA and (R)-3-hydroxyvaleryl-CoA, respectively [10]. The above mentioned reactions are summarized in Figure 1.5.

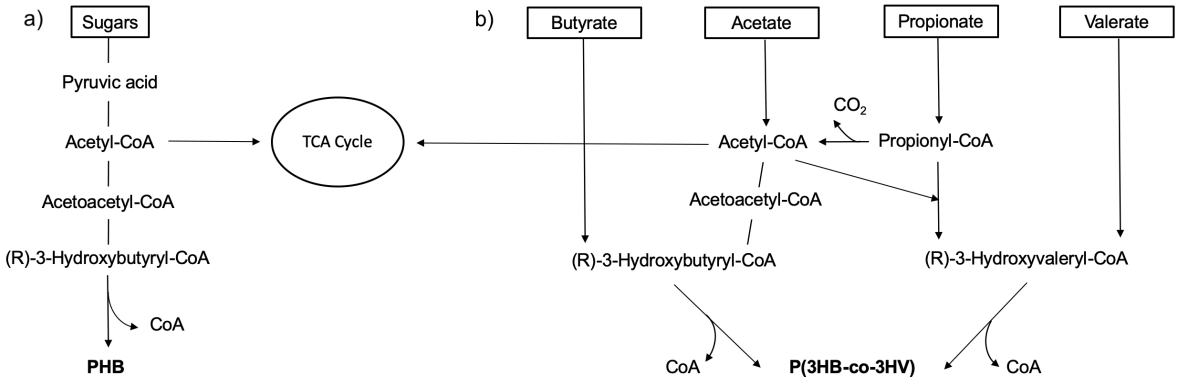


Figure 1.5: Metabolic pathways involved in polyhydroxyalkanoate (PHA) synthesis from sugars, through glycolysis (a) and from fatty acids, directly (b). Adapted from [10].

Longer-chain fatty acids can be broken down by β-oxidation, which acts in a cyclic manner, removing a two-carbon acetyl-CoA group from the fatty acyl chain each cycle. [61] Apart from acetyl-CoA, which, as mentioned above, is one of the main building blocks for PHB synthesis, the different intermediates in the cycle may also be converted to the corresponding (R)-3-hydroxyacyl-CoA and, subsequently, originate mcl-PHA (or scl-PHA once the fatty acyl chain is short enough) [59]. Reversely, the *de novo* synthesis pathway builds up fatty acids from acetyl-CoA by adding a new acetyl-CoA group in each cy-

cle and also provides intermediates suitable for conversion into (R)-3-hydroxyacyl-CoA monomers (see Figure 1.6 [61].

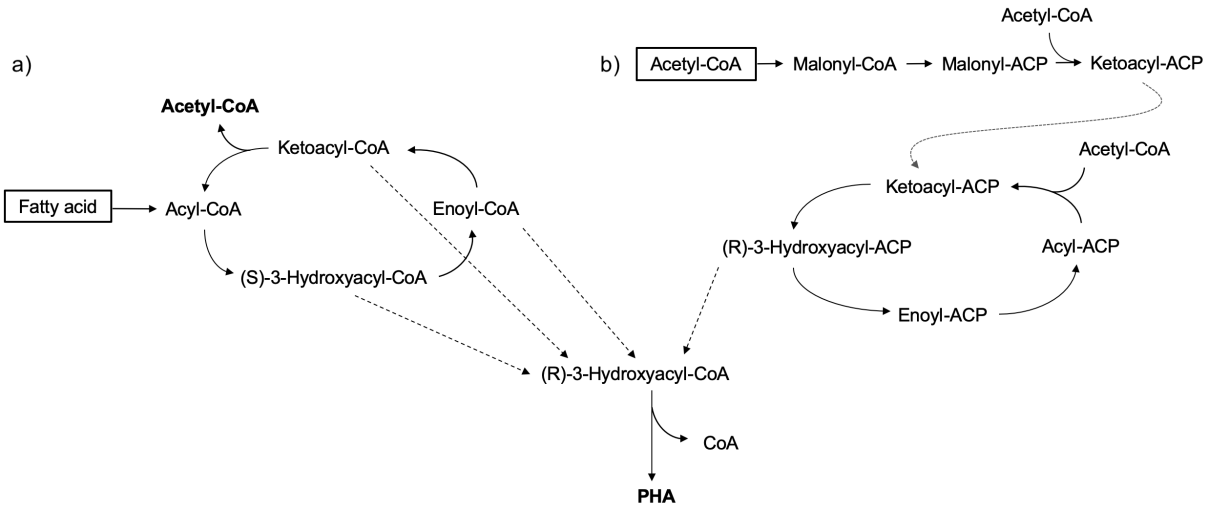


Figure 1.6: Metabolic pathways involved in polyhydroxyalkanoate (PHA) synthesis using intermediates from β-oxidation (a) and from *de novo* synthesis (b) as precursors. Adapted from [61].

However, when a mixture of VFAs - and not a single substrate - is used, intracellular regulation of PHA synthesis is significantly more complex, since not one but several of the above mentioned pathways are activated [62]. Many experiments have attempted to provide a correlation between the profile of VFAs in the feed stream and the polymer final composition, and some metabolic models have even been developed [62, 63]. For example, the balance of (3HB) and (3HV) in the PHA copolymer can be influenced by even versus odd number of carbon atoms in the VFA feed [64]. Takabatake *et al.* achieved increasing molar ratios of (3HV) in a P(3HB-co-HV) copolymer, from 3% to 84%, when the propionate fraction in a mixture of propionate and acetate was increased from 0% to 100% [65]. Furthermore, Albuquerque *et.al* studied mixtures of acetate, propionate, butyrate and valerate in fractions of 30/20/30/20 and 60/15/20/05 and obtained P(3HB-co-HV) copolymers with around 70:30 and 85:15 HB:HV ratio, respectively [66].

1.3.4 Applications

Over the last decade, PHAs have witnessed a significant development at the industrial and commercial level, thanks to a number of emerging partnerships and collaborations between the top players in the market [67]. Danimer Scientific [28], for example, acquired the patented PHA technology from Procter & Gamble in 2007 [68] and is also collaborating with the University of Georgia since 2015 [69], strongly focusing on innovations through extensive R&D activities. In 2018, its Nodax-brand polyhydroxyalkanoate biopolymer was used to create the first fully biodegradable plastic straw [70] and is currently working with PepsiCo to develop a chip bag made of the same polymer [71].

These latest advances in PHA production at a large scale have opened up countless opportunities regarding their applications. Nowadays, PHAs aim to replace conventional fuel-based polymers in a variety of fields. Regarding packaging and coating applications, PHAs can be used to produce a variety of products, including shopping bags, collection bags for solid wastes, single-use containers for hygiene and cosmetic products, agricultural films, bottles, cups and materials for food packaging [72]. These applications fundamentally overlap those of biodegradable plastics mentioned in Section 1.2.2.

On the other hand, there is another unique and appealing characteristic associated to PHAs: their biocompatibility. The 3-hydroxyacids which compose the polymer are naturally found in animals and

it has already been shown that they do not produce toxic or immunological responses in the human body. This has made them suitable candidates for the development of implants, sutures and scaffolds for tissue engineering and tissue repair [73]. A number of promising studies have been reported, showing the potential of PHAs to outperform other polymeric materials as three-dimensional scaffolds to support cell growth for tissue engineering of blood vessels, heart valves, bone [74], cartilage [75], nerve and skin. Furthermore, different copolymers have been used to develop materials for wound management, such as sutures, implants, stents, patches and pins [76].

Also making use of both biodegradability and biocompatibility, the pharmaceutical sector is working on the development of micro and nanospheres of PHA that can be utilized as as delivery systems for hydrophobic drugs. The principle is that the active compound is released once the polymer coating starts to degrade [77]. The hydrophobic nature of PHA can be modified to encapsulate hydrophilic drugs. Furthermore, PHAs in the form of nano-beads have been used together with proteins that intrinsically bind to the PHA granules to immobilize recombinant proteins of interest and design functional nanocarriers for a wide range of applications, such as drug screening and diagnostic tools [78].

Apart from the above mentioned applications, PHAs show promise in endless other fields: as biofuels, raw materials for the production of paints, fiber materials, paper waterproof coatings, among others [79]. However, for many of these opportunities to become reality, the high costs associated to PHA production must be tackled. In this context, production with MMCs will now be introduced.

1.4 Production of Polyhydroxyalkanoates with Mixed Microbial Cultures

1.4.1 Substrates

Over the last two decades, numerous studies have focused on producing PHA using waste carbon sources from the agro-food and industrial sectors, as well as from urban waste treatment plants. Apart from decreasing the production costs, using waste streams as substrates also adds the advantage of significantly reducing their polluting load [10].

These complex feedstocks are characterized by having a high organic fraction and comprise, among others, cheese whey effluents, fruit and vegetable solid wastes, olive oil mill wastewater, crude glycerol and the organic fraction of municipal solid waste leachate. Lignocellulosic materials (e.g. wood, agricultural, or forestry wastes) are available in the highest quantity. Lately, there has been an increasing worldwide effort to create biorefinery plants where lignocellulosic wastes can be converted to starting materials not only for production of biopolymers but also of bioethanol and a range of fine chemicals [80, 81, 82]. To avoid the logistics for collection and transportation of these feedstocks, these facilities are, preferably, integrated into existing production lines [38].

Another factor to be considered is the availability of the waste stream throughout the year. In Europe and North America, surplus whey from the dairy industry is continuously available in large quantities, whereas in other parts of the world, the availability of lignocellulosic materials is compromised when no harvest takes place (off-season). Therefore, storage strategies are necessary, usually with the conversion of perishable components into stable intermediates, like lactic acid. Lastly, the substrate's composition should have minimal variability in order to ensure that the final quality and characteristics of the polymer are uniform and reproducible [81].

1.4.2 3-Stage Production Process

Processes for PHA production involving MMCs are commonly operated in three-stages, comprising: (1) an acidogenic fermentation of the organic carbon substrate to produce VFAs, which are used as precursors for PHA biosynthesis; (2) a culture selection stage, where the culture is enriched with PHA-producing organisms by applying a selective pressure method for PHA storage; and (3) a PHA accumulation stage where the selected culture is fed with an excess of VFAs in order to exploit its PHA storage capacity to its maximum [83]. Figure 1.7 summarizes the typical process layout.

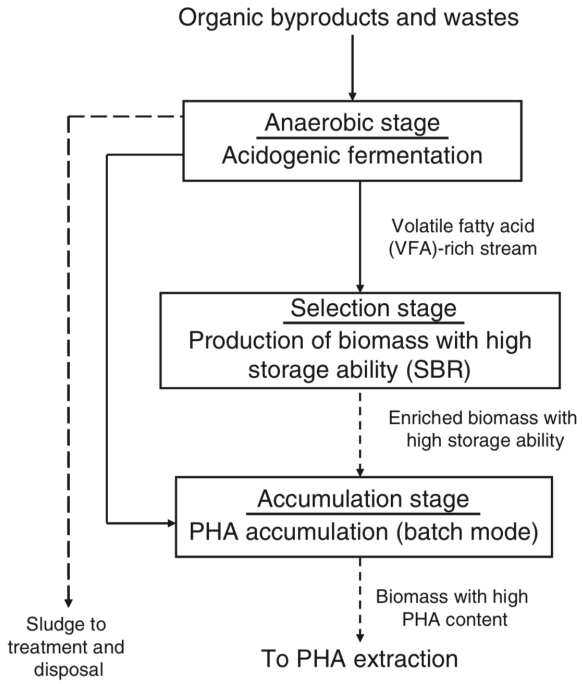


Figure 1.7: Schematic representation of the 3-stage process for PHA production with mixed microbial cultures. The dashed lines indicate sludge displacement and the solid lines refer to the movement of the VFA-rich mixture. SBR stands for Selection Batch Reactor. Adapted from [10]

Acidogenic Fermentation

Anaerobic digestion (AD) has been used for the treatment of a wide variety of organic waste streams for decades now, as it allows for the conversion of complex organic matter into simple carbon compounds, such as carbon dioxide and methane [84, 85]. This conversion is usually carried out by a consortium of both facultative and strict anaerobic microorganisms, through a sequence of biochemical processes. The first step is an enzymatic hydrolysis, in which initial proteins and polysaccharides are broken down to monomeric sugars, amino acids, long chain fatty acids and alcohols. Acidogenic bacteria are then capable of carrying out the fermentation of these simple organic compounds, which leads to the production of volatile fatty acids, hydrogen and carbon dioxide, in a step called acidogenesis. These products can, however, be further oxidized to acetate, hydrogen, and carbon dioxide, which is known as acetogenesis. Lastly, a consortium of methanogenic bacteria are responsible for producing methane, either by decarboxylation of acetate or by reduction of H_2/CO_2 [86].

The VFAs commonly used for PHA production with MMC can, therefore, be produced from the acidogenic fermentation of organic waste streams [87]. Hydrolysis may or may not take place, depending

on the type of substrate. It is preponderant, however, that methanogenesis - and often acetogenesis - are prevented, since it is the intermediate byproducts that are of interest in this case. This can be done by selecting specific operating conditions that inhibit the activity of methanogenic and/or acetogenic microorganisms, such as low solids retention time (SRT), low temperature, low pH, or a combination of these different approaches [10].

This fermentation is often the first step in the production process of PHA with MMCs. Manipulating the operating conditions of the corresponding bioreactor can lead to a variety of VFA profiles and concentrations from the same original waste stream, which in turn can originate polymers with different monomer composition, as was just mentioned in the previous section [88]. The different parameters that can be tuned in order to achieve a specific VFA composition include hydraulic retention time (HRT), SRT, pH, temperature, organic loading rate (OLR) and reactor configuration [10]. For example, the concentration of propionate was found to increase with both pH (within an acidic range) and HRT for paper mill and cheese whey effluents [89], while butyrate and valerate were preferably produced at lower pH when using sugarcane molasses [90].

Lastly, it is important to mention that when the organic waste stream chosen as carbon source for PHA production is already mainly composed of VFAs, this first acidogenic step is unnecessary. Examples of such wastes include, for example, methanol in pulp and paper mill foul condensate or a mixture of ethanol, glycerol, esters and fatty acids in biodiesel wastewater [91]. In these cases, the substrate is directly used for both MMC selection and accumulation stages, which will now be approached.

Selection

As mentioned in Section 1.3.2, if MMCs are to be used for PHA production, a selective pressure must be applied so that the culture is enriched with PHA-storing organisms. The presence of microorganisms with low storage capacities would affect both the productivity of the process and the downstream stage, increasing final production costs. [10] It is important to emphasize that the aim of this step is to achieve a population with high and stable storage capacity, rather than to maximize the PHA cell content [59]. Indeed, the latter is achieved during the final accumulation step, which will be approached in the following section.

The selection usually takes place in a selection batch reactor (SBR) which operates in aerobic conditions and is fed with the VFA-rich stream produced during the acidogenic fermentation. Generally, this reactor is operated under short periods of carbon substrate abundance (feast phase), alternated with rather long periods of carbon unavailability (famine phase). During the latter phase, PHA-producing microorganisms are able to use their internal carbon reserves and external nitrogen for growth and maintenance, whereas organisms with no storage capacity are unable to do so and will, eventually, be eliminated from the reactor [92]. This type of enrichment technique is commonly known as Aerobic Dynamic Feeding (ADF) and generally allows for the selection of cultures highly enriched in PHA-storing organisms [93].

Moreover, during starvation periods, the activity of cells is reduced to a minimum and the rates of enzyme synthesis and RNA transcription are significantly decreased [37]. This internal limitation means that growth is inhibited in the next feast period and cell physiological adaptation is needed to reach maximum growth rates. Nevertheless, PHA synthase remains active and substrate uptake is, therefore, mainly driven toward polymer storage during these adaptation phase [93]. This faster response provides PHA-producing microorganisms with a new competitive advantage [37].

A different approach for culture enrichment is based on anaerobic/aerobic cycles. In fact, PHA storage by MMC was first observed in Enhanced Biological Phosphate Removal (EBPR) systems, in wastewater treatment plants (WWTP), where polyphosphate-accumulating organisms (PAOs) and GAOs were described to accumulate PHA during the anaerobic phase [94]. These dynamic conditions of oxygen availability follow a similar principle to ADF: the external carbon source is only available during the anaerobic phase, PAOs and GAOs store it as PHA, while glycogen, a second storage polymer, is consumed; in the subsequent aerobic stage, when oxygen is available for oxidative phosphorylation, PHA is used for cell growth, maintenance, and glycogen pool replenishment and a selective advantage is provided to PHA-producing microorganisms [10]. Although ADF has presented itself as the preferred approach when only PHA production is intended, there is still potential in optimizing these EBPR systems for both wastewater treatment and resource recovery [95].

Several operational parameters regulate culture selection in the SBR reactor. One of the key factors is the feast to famine (F/F) length ratio: the famine phase must be long enough to ensure internal limitation, otherwise the culture will be better fit to grow when supplied with external carbon source and will therefore accumulate less PHA, hindering the selective advantage of PHA-producing organisms [96, 92]. A low F/F ratio is usually achieved by implementing low OLR or long cycle lengths. Substrate concentration in the reactor is another important parameter: substrate uptake and polymer production rates increase with substrate concentration up to a certain maximum value and after that decrease due to inhibition by the substrate. This inhibition can extend to the feast phase and, consequently, alter the pre-established F/F ratio [93].

The SRT also plays an important role. Since the effective specific growth rate of a culture decreases with the increase in SRT, keeping this parameter at higher levels is expected to inhibit growth and, therefore, promote PHA storage. On the other hand, a shorter SRT leads to a stronger weeding out effect, which favours PHA-producing organisms by enhancing their growth advantage. In fact, contradicting results can be found in the literature [97, 98, 99], suggesting that further studies are required to understand the effects of SRT on selection processes. Moreover, temperature and pH can also impact the selection performance, although the effects depend mainly on the microbial composition of the culture and not much can be inferred but that pH is usually kept above 7 [99, 100].

Another important factor is the carbon to nitrogen (C/N) ratio and the corresponding nutrient feeding strategy. Generally, studies have found that the polymer storage ability of the biomass can be improved under dynamic conditions with nitrogen deficiency, when compared to a nutrient excess scenario [101, 102]. However, since carbon is preferably directed towards polymer storage instead of growth, this means that the overall volumetric productivity is compromised. Regading the feeding strategy, nitrogen can either be fed together with the carbon substrate during the feast period, or it can be uncoupled and delivered during the famine stage. This latter strategy was proved to select a more efficient PHA storing culture which does not depend on the F/F ratio. This means that higher OLRs can be applied to the system, which allows for higher volumetric productivities without compromising the efficiency of the enrichment step [92, 103].

To sum up, the main goal of the selection stage is to ensure that the MMC is properly enriched with PHA-producing organisms. Furthermore, since this culture is to be used in the subsequent production stage, high cell densities would allow for a more concentrated biomass inoculum. Therefore, the ultimate goal of the SBR operation is to select microbial cultures capable of high growth rates without compromising their high storage capacity [96].

Accumulation

The final step in the 3-stage production process is accumulation and its main goal is to exploit the culture's PHA production capacity to its full potential. Thus, the enriched MMC that is obtained after selection is fed with the VFA-rich stream resulting from acidogenic fermentation, usually in the absence of nutrients, so that the carbon is mostly channelled towards PHA storage rather than growth [92]. In fact, biomass yield is generally achieved during the previous selection step and the accumulation's goal is mainly to increase the PHA content in this biomass.

This operation is commonly carried out in fed-batch mode, with the carbon substrate being fed pulse wise. The length of each pulse can be monitored indirectly through dissolved oxygen (DO) measurements: when an external carbon source is available, DO decreases due to substrate consumption [104]; once the carbon source is depleted, a sudden increase in DO is observed and a new pulse of substrate is fed to the accumulation reactor. This procedure is repeated until there is no DO response [83], which is when the culture is believed to have reached its maximum storage capacity. Ideally, no PHA consumption takes place between pulses and the biomass is exposed to conditions of extended feast, so to say [64]. It is worth mentioning that several continuous feeding strategies have been studied and show potential when regarding process scale-up [66, 105, 106, 107].

However, the overall PHA output is still significantly low for MMCs, mostly because of the cell densities attained in the selection phase, and several studies have now been reported on which higher PHA volumetric productivities were achieved with targeted levels of nutrient limitation, suggesting that it might be possible to promote growth of the PHA-storing biomass during the accumulation stage without it overtaking storage response [108, 109]. This is very promising for PHA production using waste feedstocks as carbon sources, since many fermentable agro-industrial organic wastes are far from nutrient-limiting.

As in the enrichment step, several operational parameters have an impact on PHA accumulation. The influent substrate concentration is usually that which is fed to the SBR during the feast phase, just like the pH and temperature tend to be kept at the same values, since the culture is fundamentally the same [93, 92, 83]. However, it is the success of the selection that influences the outcome of PHA accumulation the most [64]. A culture with a considerable cell density, enriched with PHA-producing microorganisms that are well adapted to the carbon feedstock is needed in order for the accumulation stage to be a success. Table 1.1 summarizes the achievements of some pilot-plant facilities producing PHA through this 3-stage production process worldwide.

1.4.3 Downstream Process

Because PHAs are intracellular compounds, downstream processing constitutes a considerable cost factor in the polymer production process, especially when mixed cultures and waste feedstocks are involved, not only because PHA has to be isolated from a more complex matrix but also because the polymer is considerably diluted as a result of the low cell densities attained during production with MMCs. The PHA content in the biomass also affects the efficiency of polymer recovery: generally, less digesting/extracting agent is needed to retrieve PHA granules from cells with higher polymer content [115].

Nevertheless, recovery of PHA is a complicated process that normally requires more than a single unit operation: firstly, right after the accumulation step takes place, the biomass must be separated from the broth (e.g. through centrifugation, filtration, sedimentation); afterwards, the biomass can be pre-treated to promote the permeability of the cells (e.g. by heating, freezing, adding salts) and is then

Table 1.1: Examples of pilot-scale projects for the production of polyhydroxyalkanoates (PHA) from different waste feedstocks and their main features. The productivity of the process can be evaluated by the substrate to product yield and intracellular PHA content (%w/w), although standardization of these parameters is still not fully guaranteed and, thus, comparison between studies is often not straightforward. * Yield calculated on a COD basis.

Pilot Plant, Location	Feedstock	Origin of MMC and Enrichment Strategy	$Y_{P/S}$ (g/g)	PHA Content (%)	HB:HV Ratio (%w/w)	Ref.
Nagpur, India	Pre-fermented milk and ice cream processing wastewater	Activated sludge	0.425*	39-43	-	[110]
Eslov, Sweden	Beet processed water, 38% in VFA	PHA producing MMC from pre-fermented effluent of Procordia Foods	-	60	85:15	[111]
Brussels North WWTP, Belgium	Pre-hydrolyzed and fermented WWTP sludge	Sludge under aerobic feast famine	0.25-0.38	27-38	66-74:26-34	[112]
Leerwarden WWTP, Friesland, Netherlands	Fermented residuals from green-house tomato production	Sludge under anoxic feast/ aerobic famine	0.30-0.39	34-42	51-58:42:49	[113]
Mars company, Veghel, Netherlands	Fermented wastewater from a candy bar factory	Activated sludge under aerobic feast famine with inhibitor of nitrification	0.30	70-76	84:16	[114]

subjected to an extraction process, in which either the non-PHA cellular material or the PHA are solubilized; lastly, PHA is isolated and purified [116].

Regading the extraction step, the most commonly used technique is solvent extraction, given its simplicity and the high purity it allows for. However, this method makes use of significant amounts of hazardous solvents and energy, which have a negative impact on both the total production cost and the environmental footprint of the process, compromising the whole sustainability motivation. Thus, alternative processes are emerging, from chemical/enzymatic digestion of non-PHA cellular biomass [117, 118], to extraction with supercritical fluids [119], which have a low toxic nature and cost. Lately, aqueous-two phase systems have also been described and show promise considering both economical and environmental factors [120].

Lastly, choosing an extraction method for a particular process must not only account for production costs and environmental impact but also take into consideration the intended application for the polymer. For example, if the extracted PHA is to be used for the fabrication of a medical product, high purity rather than cost is the key factor, whereas if the final product is a single-use disposable item, then high purity is no longer mandatory and cost will probably take over as the determinant factor [56].

1.4.4 Challenges and Future Perspectives

PHA production using waste-based feedstocks and enriched mixed cultures is now well established. Apart from reducing operational costs by not requiring sterile conditions nor costly substrates, MMC-based processes offer the possibility of using a wide range of waste streams, reducing their pollutant load and allowing for a possible process integration in current wastewater treatment plants or industrial facilities. In fact, wastewaster, crude glycerol and whey are gaining attention as potential carbon sources, given their adequate and constant supply. On the other hand, the combination of two or more feedstocks could avoid the requirement for additional synthetic nutrient and/or mineral feeding in the process.

The three-stage process has been demonstrated at both laboratory and pilot scale and PHA contents comparable to those of pure culture processes have been demonstrated. However, the low cell densities and consequent volumetric productivities attained with MMCs are the main bottleneck of the process, along with some challenges in the cost and environmental impact of the downstream stage. One potential solution to tackle the cell density issue could be based on the use of membrane bioreactors (submerged or cell recycling), that allow for an effective retention of biomass inside the reactor. Regarding PHA recovery, there is an urgent call for a large scale integration of novel downstream strategies that avoid toxic organic solvents, many of which have already been successfully applied in the laboratory, such as the aqueous two-phase system or enzymatic methods.

Further research is required to understand the effect of key controlling factors in the accumulation process, such as the characteristics of the selected culture, the feeding strategy (e.g. continuous vs pulse, nutrient excess vs limitation), and the operational parameters of the reactor (e.g. pH and temperature). In this context, online monitoring techniques are also necessary for process optimization. Moreover, for the development of polymers with specific chemomechanical properties, the metabolism of mixed cultures and the way different substrates translate into certain polymer compositions must be better understood. Only when both the operational conditions and substrate conversion are finely tuned will the tailored and reproducible synthesis of PHA with MMCs and waste streams become a reality.

Global environmental awareness, continuous efforts in different scientific areas and cooperation between the academia and industry are needed in order to solve the current challenges and take PHA production with mixed cultures to an industrial and commercial level.

1.5 Spectroscopy for Monitoring PHA Production

Nowadays, real-time monitoring of bioreactors is considered essential for a reliable and robust process control and, consequently, for improved efficiency, productivity and reproducibility [121]. For instance, measurements of cell growth, substrate consumption, and product concentration are of utter importance in bioprocess operations, such as PHA production. Traditionally, off-line analytical methods like gas chromatography (GC) and high-performance liquid chromatography (HPLC) are carried out to obtain such measurements [122]. However, these are destructive, labor-intensive and time-consuming techniques, meaning the results are usually too late to be of any use for process control [123] Thus, there has been an increasing effort to develop more efficient monitoring tools.

An ideal sensor for real-time monitoring of bioprocesses should be rapid, sensitive, non-destructive and robust. Furthermore, the system should preferably generate multi-analyte data, require minimal sampling, and endure the complex bioreactor environment.

A possible set of candidates are on-line sensors, such as optical sensors, chemosensors and biosensors, which have shown improved performance and reliability [124]. However, they are generally costly, require high-maintenance and provide single-property analysis, limiting their applicability. Spectroscopy, on the other hand, has been gaining attention for on-line, in-situ monitoring of bioreactors, given the recent advances in the development of robust spectrometers, high-quality probes and optical fibers, and the possibility of coupling with chemometrics tools for process modelling and multi-analyte quantification [121].

Among the different spectroscopy methods available for monitoring PHA production, vibrational spectroscopy has received the most attention. In this context, Raman scattering and near-infrared (NIR) spectroscopy will be approached with greater detail in the subsequent sections, as they represent the scope of the present thesis.

1.5.1 Raman Spectroscopy

Fundamentals

When electromagnetic radiation interacts with a molecule, the photons can either be absorbed or scattered. For absorption to take place, the frequency of the incident light must correspond to the energy gap between the ground state of the molecule and an excited state. Therefore, in absorption techniques, such as infrared spectroscopy, a wide range of frequencies is directed onto the sample in order to find out which ones are missing after interaction. On the other hand, scattering-based vibrational spectroscopy, for which Raman is the best example, does not require the photons to exactly match the energy difference between a pair of states, but instead, the incident light interacts with the molecule and is scattered from it. In this case, a single frequency radiation is used and the frequency difference between the incident and the scattered light is detected [125].

In scattering processes, this interaction between light and matter is translated in a distortion of the electron cloud around the nuclei that brings the molecule to a so-called 'virtual state'. This 'complex' is highly unstable and the energy is quickly released as scattered radiation. If only electron cloud distortion is induced, light is scattered with the same frequency, since electrons are light and cause unsignificant frequency changes. Most photons scatter through this elastic phenomenom, known as Rayleigh scattering. However, if nuclear motion is involved, there is energy transfer between the light source and the molecule and a shift in the frequency of the scattered radiation can be observed. This inelastic scattering of monochromatic light is known as Raman scattering and is the basis of Raman spectroscopy, since the frequency shift depends on the chemical bonds of the molecules and can, therefore, provide qualitative and quantitative information about the samples. However, Raman scattering is considerably less likely than Rayleigh's, with an intensity 10^{-6}-10^{-8} that of the incident light, making this spectroscopy method an inherently weak process [125].

Figure 1.8 illustrates the Rayleigh scattering phenomenom, along with the two types of Raman scattering: Stokes and anti-Stokes. A Stokes shift takes place when the molecule is initially in the ground state and energy is transferred from the incident photon to the molecule, promoting it to a higher energy state and increasing the wavelength of the former. An anti-Stokes shift, on the other hand, takes place when the molecule is already in an excited vibrational level due to thermal energy and is brought back to the ground state after interaction with the electromagnetic radiation. In this case, the photon reaches the detector with a shorter wavelength than the light source. In the end, Raman spectroscopy measures the difference between the initial and the final states by subtracting the energy of the scattered photon from

that of the incident beam. The final vibrational spectra can provide data about the composition, chemical environment, and structure of a given sample [126].

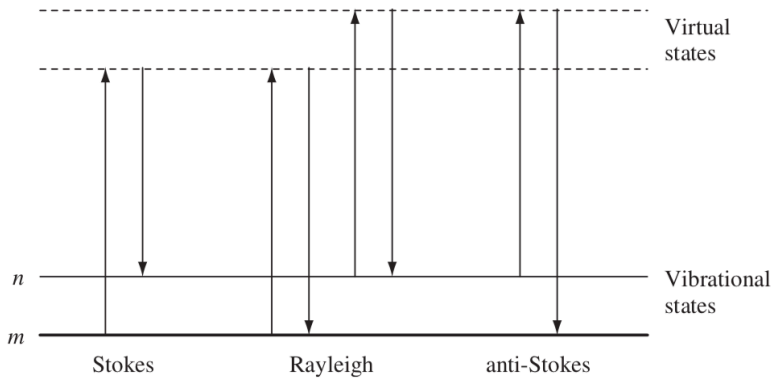


Figure 1.8: Diagram of the Rayleigh, Stokes and anti-Stokes Raman scattering processes. The lowest energy vibrational state is represented as m and n represents higher energy levels. Both the incident laser energy (upward arrows) and the scattered energy (downward arrows) have much larger energies than the energy of a vibration. The dashed lines indicate the "virtual states."

In a typical Raman experiment, a laser is used to irradiate the sample with monochromatic radiation. Different choices of Raman lasers are available, which can be broadly classified into three different categories depending on their wavelength: ultraviolet (UV) lasers (<360 nm), visible lasers (450-660 nm), and near-infrared (NIR) lasers (785–1064 nm). Apart from the laser, any Raman spectromoter will have to include a filter to eliminate Rayleigh scattering, so that only Raman scattering is detected [121].

Advantages and Disadvantages

One of the main advantages of Raman spectroscopy is the fact that it can be applied to basically every type of sample (i.e. solids, liquids, gases, slurries, gels, etc). The method is non-destructive, little to no sample preparation is needed and an immersion probe can even be used to directly acquire the spectra (e.g. by inserting the probe inside the bioreactor) [121]. Regarding the information it provides, Raman spectroscopy can be used for multi-analyte analysis and is best at detecting the symmetric vibrations of non-polar groups, since these cause the largest changes in the polarizability of the molecule, and hence, the greatest scattering, in contrast with IR spectroscopy, which is more appropriate for identifying assymmetric vibrations of polar groups [126]. In fact, Raman and IR spectroscopy can be seen as complementary techniques and usually a combination of both is needed in order to get a full view of the vibrational modes of a molecule. Furthermore, Raman methods are compatible with aqueous solutions, since the spectra are not sensitive to water, while IR, on the contrary, is strongly limited in such cases due to the excessive absorptivity of water [127].

The most obvious disadvantage of Raman spectroscopy is the weak signal it provides. This issue can be tackled by using shorter wavelengths, since Raman scattering intensity varies as the fourth power of the inverse of the wavelength [127]. However, decreasing the wavelength entails another major problem in Raman spectroscopy: a high fluorescence background that can easily overcome the weak Raman signal, rendering the spectra useless. Thus, selection of laser wavelength requires a delicate balance between maximizing the signal and minimizing fluorescence interference [121]. Nevertheless, signal intensity can also be improved by the laser flux power density. In fact, most modern Raman equipments

are coupled with microscopy tools, allowing for the delivery of very high power densities to very small samples [125].

Applications for PHA Production Process

Most studies using Raman as a monitoring tool for PHA production have focused on intracellular polymer quantification, composition and degree of crystallinity. De Gelder *et al.* compared the spectra of PHB-producing *C. necator* DSM 428 and its PHB-negative mutant DSM 541 during growth and were able to locate the most prominent contributions of PHB to a bacterial Raman spectrum. In particular, a peak at around 1734 cm^{-1} was considered suitable to follow PHB synthesis and consumption during growth. The intensity of the Raman band was proven to be proportional to PHB concentration and, therefore, the authors present Raman spectroscopy as a fast and efficient tool with potential for monitoring and quantifying the intracellular PHB content in biomass [128]. Samek *et al.* also identified a peak around the same wavenumber (1736 cm^{-1}) in *C. necator* H16 and carried out a calibration model that can be used as a fast and reliable determination of intracellular PHB content, replacing time-consuming GC analysis [129].

Izumi and Temperini used commercially available copolymers of P(3HB-co-HV) with HV contents between 0 and 12% and were able to locate several Raman bands for amorphous/crystalline domains and to quantify the molar fraction of 3HV in the samples through the ratio of the intensity of the bands at 2938 cm^{-1}/1740 cm^{-1} and 1354 cm^{-1}/1740 cm^{-1} for polyester solutions and molten polyester films, respectively [130]. Similarly, Jost *et al.* tempered P(3HB-co-HV) samples in order to create different degrees of crystallinity (0.69-0.86) and developed a model that considers the ratio between the sensitive band at 2999 cm^{-1} and the internal standard at 1059 cm^{-1} to estimate this physical parameter. Furthermore, the study also presented a model for calculation of the 3HV content (0-12%) in copolymer films using the ratio between the sensitive band at 844 cm^{-1} and the internal standard at 1101 $cm-1$ [131]. No models have yet been developed for estimation of 3HV content and degree of crystallinity of intracellular polymers (still inside the cells).

Apart from estimating PHA content and composition, there is potential for using Raman spectroscopy to monitor cell growth. Petrosino reported the development of a preliminary model that follows the total suspended solids (TSS) and volatile suspended solids (VSS) concentration in an SBR reactor belonging to a pilot-scale 3-stage process for PHA production [132]. Despite this and the other above-mentioned successes, the use of Raman spectroscopy for on-line monitoring of bioprocesses is still very limited, especially for production with MMCs.

1.5.2 Near-Infrared Spectroscopy

Fundamentals

Near-infrared (NIR) spectroscopy is a type of vibrational spectroscopy that employs photon energy in the wavelength range of 780 to 2,500 nm (wavenumber between 13,000 to 4,000 cm^{-1}) [121, 133]. As in other absorption techniques, it makes use of a polychromatic light source from which the sample absorbs specific frequencies that correspond to molecular vibrational transitions. With mid-infrared (MIR) radiation, these transitions are usually from ground level to the lowest excited vibrational state of a molecule, and they are known as fundamental vibrational transitions, whereas NIR radiation has

enough energy to bring the molecules to higher excitation levels. This is possible through overtones and combinations of the different fundamental bands seen in the MIR region, as will now be explained [134].

According to the simple harmonic model (A in Figure 1.9), the absorption or emission of energy can only occur when it matches the energy gap between any pair of neighboring vibration levels ($\Delta\nu$=1, where ν represents the vibrational quantum number). In practice, since most molecules in a sample at room-temperature normally occupy the lowest energy level, this means that the only significant transition is from 0 to 1, which corresponds to the above-mentioned fundamental vibrational transition promoted by MIR radiation. However, this model is not perfectly adequate for describing the vibrational motion of chemical bonds in the NIR region, due to non-ideal behaviors such as repulsion between electronic clouds when the atomic nuclei approach and molecular dissociation when the nuclei are overly displaced. On the other hand, the anharmonic model (B in Figure 1.9) is more realistic and, while it still applies some limitations to the possible energy levels of the molecules, it already predicts the occurrence of overtones (i.e. transitions with $\Delta\nu$=2 or higher) and of combination bands between two or more fundamental vibrational transitions taking place simultaneously. Moreover, it also predicts that the separation between two adjacent energy levels decreases with ν. The frequencies of both overtones and combinations are usually in the NIR region and, despite the higher complexity of the final spectra, these events can usually be assigned to specific group vibrational frequencies [135, 133].

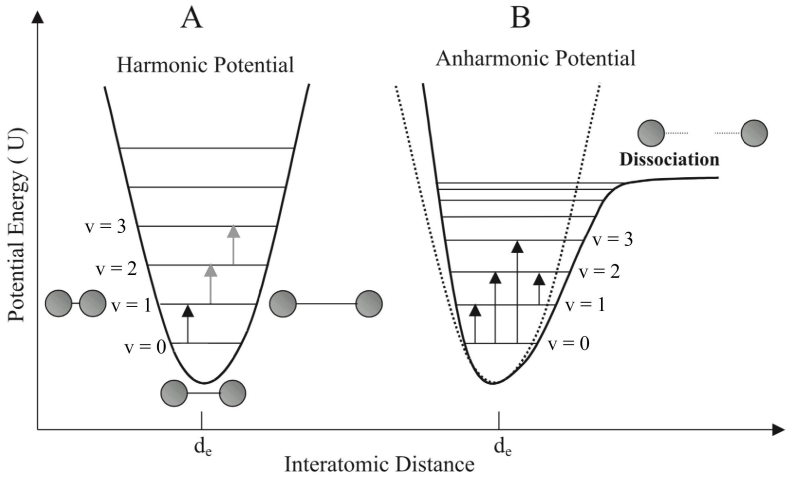


Figure 1.9: Schematic representation of the harmonic (A) and anharmonic (B) models for the potential energy of a diatomic molecule. d_e = equilibrium distance.

For a particular vibrational event to be detected through NIR (or any IR) spectroscopy, it must not only equal the frequency of the incident light but must also promote a change in the dipole moment of the molecule, which is why symmetrical streches are NIR inactive [134]. Thus, the intensity of a given NIR band is dependent on the magnitude of the dipole change during the motion of the atoms in a vibration and on its degree of anharmocity. Both phenomena are strongly present in the vibration of chemical bonds between hydrogen and a heavier atom such as carbon, nitrogen or oxygen (i.e. X-H bonds), which are the basis of most NIR spectroscopic information. The spectra of water, for instance, are characterized by the presence of strong peaks at 5,200 cm^{-1} and 7,100 cm^{-1}, which correspond to the combinations and first overtone, respectively, of the fundamental vibration of the O-H bond. Since most biological samples comprise X-H bonds, NIR spectroscopy is potentially suitable for measuring most components involved in bioprocesses [134, 135].

Advantages and Disadvantages

The bands occurring in the NIR region are between 10–100 times weaker than those obtained with MIR radiation and, therefore, no sample pretreatment is needed. Furthermore, just like Raman, NIR spectroscopy is a rapid, non-destructive technique and an immersion probe can also be used for real-time, in-situ monitoring [135, 121]. It is a flexible tool, with various measurement modes available: diffuse reflectance (for solids, powders, etc.), transflection (for slurries, semi-solids, liquids, films, emulsions, etc.) and transmitance (for clean liquids) [135, 136]. NIR spectroscopy can also be considered a multi-analyte method, since a single measurement can provide information about several different compounds present in the sample [121].

On the other hand, this non-selectivity can also be a disadvantage, because interferents can often mask the signals of interest. For instance, since water absorbs strongly in the NIR region, the applicability of NIR spectroscopy to aqueous solutions is often conditioned. Moreover, the fact that only bonds to hydrogen produce intense bands in the NIR region may also be a limitation in some situations. Lastly, as mentioned above, the presence of overtones and combinations makes NIR spectral data complex and not directly interpretable. It is the development and utilization of the mathematical techniques known as chemometrics that allows for the extraction of relevant information from the spectra, as will be introduced in Section 1.5.3. In fact, the recent developments in computation and modelling tools have made NIR a promising candidate for low-cost and real-time monitoring of bioprocesses, and of PHA production in particular [121, 137].

Applications for PHA Production Process

NIR spectroscopy has already been proven useful for detection and quantification of PHA and other relevant elements of the production process. Cruz *et al.* developed an at-line monitoring strategy based on NIR technology for the production of P(3HB) by *C. necator* using waste cooking oil as a substrate. The models were developed through statistical analysis and provided a connection between NIR spectra and the biomass, substrate and polymer concentrations in the broth [138]. The NIR predictions were compared with values obtained by offline reference methods and acceptable prediction errors were obtained. It is worth mentioning that quantification was achieved by selecting the spectral regions that rendered the models with the smallest error and not by identifying single peaks. Dai *et al.* carried out a similar experiment using MMCs and crude glycerol as a substrate. NIR data was used to follow P(3HB) content and substrate concentration in-situ and satisfactory prediction models were built for both [122].

Furthermore, Zhang *et al.* (2009) developed and validated a rapid quantification method, based on NIR spectroscopy, to estimate the concentrations of VFAs and ethanol in the effluent of an anaerobic H_2-producing bioreactor [139]. Besides, the method allowed for further distinction between individual VFAs. Similarly, Petrosino (2018) proposed a NIR-based approach to monitor the acidogenic reactor of a pilot-scale PHA production process with MMCs and apple pulp waste as a substrate [132]. A model to quantify the VFAs produced in the acidogenic fermentation stage was successfully developed, although a higher number of samples would be needed for full validation.

Overall, NIR techniques for in-situ monitoring have been widely applied to fermentations in the pharmaceutical industry and to environmental treatment processes [134]x, but there is clearly a lot of work to be developed in the context of PHA production, with very few papers on the topic to date. However, the potential of this method is undeniable and further research is needed in order to establish it as an effective monitoring tool to assist bioreactor control during PHA production processes.

1.5.3 Chemometrics for Spectral Analysis

Chemometrics is the science of extracting information from measurements made on a chemical system through the use of mathematical and statistical tools. These measurements can take different forms, from chromatographic (e.g. HPLC, GC) to physical (e.g. temperature, pressure) and spectroscopic (NIR, Raman). The resulting data serves as input for the development of empirical or semi-empirical models, which ultimately aim to make future predictions of the system's properties based on new measurements. Chemometrics focuses both on qualitative analysis, by creating classification models, and on quantitative analysis, through the construction of regression models [140, 141].

Since today's laboratory instruments produce large amounts of data, chemical systems are typically multivariate. For spectroscopy, in particular, each wavelength measured constitutes a different variable and considering that a common spectrometer usually detects hundreds to thousands of wavelengths per run, this means that mathematics and statistics are preponderant for establishing relationships between these variables and extracting relevant information from the system [142].

There is a generalized workflow for multivariate analysis (see Figure 1.10). After proper sample preparation and data acquisition, preprocessing is employed to enhance the signal and reduce the noise in the data (i.e. anything besides the information of interest). Secondly, a qualitative or quantitative method is carried out and a preliminary model arises. The main difference between these two approaches is that a quantitative approach requires data from a well-established analytical tool to use as reference for calibration. Lastly, internal and external validation are essential to ensure the robustness of the final model, which may then be used as a prediction tool for the system in question.

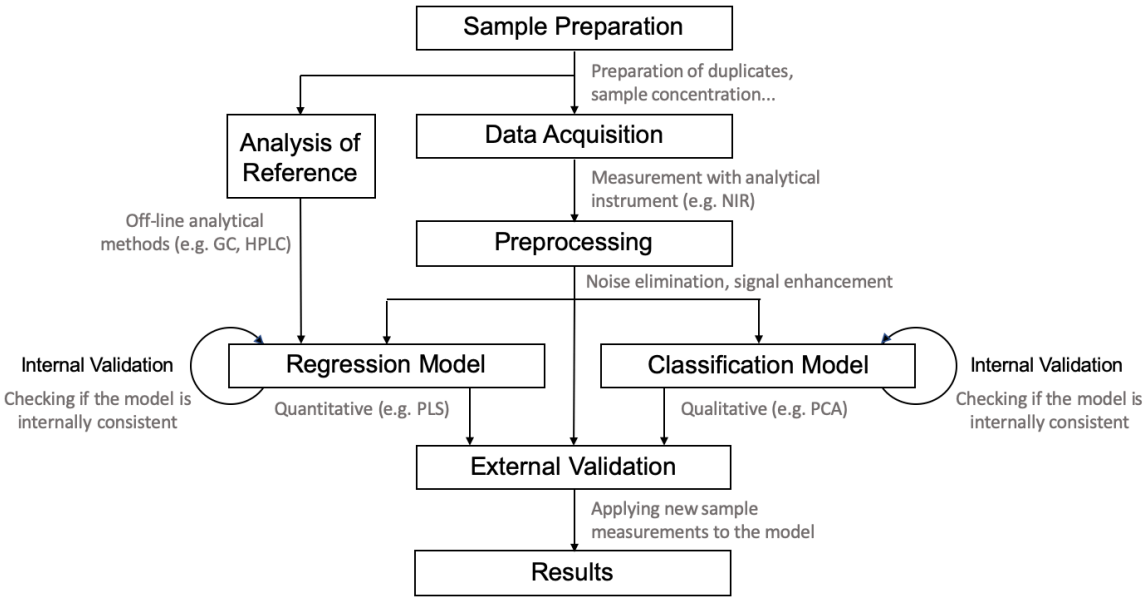


Figure 1.10: Multivariate Analysis Workflow for Model Development. PCA - Principal Component Analysis; PLS - Partial Least Squares.

Mathematical Preprocessing

Mathematical preprocessing is essential in multivariate analysis, since there wil be plenty of variation in the data that does not constitute analytical information. Thus, this step aims to eliminate unwanted signals such as fluorescence, detector noise, calibration errors, cosmic rays, laser power fluctuations, sig-

nals from cell media or glass substrate etc. Furthermore, preprocessing can enhance subtle differences between different samples [143]. Although it is common for scientists to try out different preprocessing combinations and assess which one yields the best performance, prior knowledge and understading is preferable to random trials, to avoid accidentally masking the signal of interest.

For Raman spectra, smoothing is often useful to reduce random noise and improve the visual aspect of the spectra, although it might inadvertently remove important information at an early stage. In this context, Savitzky-Golay (SG) filtering is one of the commonly used smoothing techniques [144]. Moreover, background correction/baseline removal is often carried out to eliminate the effects of broad bands or low frequency components. This is particularly relevant for Raman spectra with a fluorescence background and is usually achieved through polynomial fitting methods [145, 146]. Normalization techniques are also essential to cope with disparities in intensity. The Standard Normal Variate (SNV) normalization is a simple, but effective procedure for making spectra comparable. In Raman spectra, these intensity shifts are usually induced by laser fluctuations [145].

The most common preprocessing strategies for NIR spectra come down to spectral normalization and differentiation. Normalization is achieved through scatter-correction methods, such as the above-mentioned SNV, which aim to compensate for physical variabilities between samples or path-length variations. Differentiation, on the other hand, allows for signal enhancement by emphasizing higher frequencies in detriment of lower ones. However, this also leads to accentuated noise, which is why the Savitzky-Golay algorithm is used to simultaneously smooth the data as it takes the derivative. In some cases, it might be necessary to apply both types of pretreatment techniques, consecutively, to optimize the results.

Laslty, mean-centering is always employed, for both Raman and NIR spectra, to ensure that all results can be interpreted as variation around the mean. Further details regarding the description and algorithms of the different preprocessing tools mentioned above can be found in Subsection 3.4.1.

Qualitative Analysis: Principal Component Analysis

Qualitative analysis is usually carried out by unsupervised methods, from which Principal Component Analysis (PCA) is a good example. These methods dismiss any labelling and attempt to find hidden structures in the unlabelled data. They are often used as precursors to supervised methods when working on huge data sets [147] and can help identify outliers within them [141].

PCA is mainly used to visualize and describe the data. The method is based on projecting a multivariate data matrix (X) onto a low dimensional space, while retaining the most relevant information for further analysis. It basically reduces a set of variables into a smaller set of orthogonal, and therefore independent, principal components (PCs) in the directions of maximal variation. Each PC can be represented as a linear combination of the original variables [147]. Basically, the first PC (PC_1) will capture the maximum variance in the data and only then will the second principal component (PC_2) be determined, representing the largest residual variance along a direction orthogonal to PC_1, and so forth. From a certain PC on, the algorithm will begin to model noise and, therefore, identifying the optimal number of PCs (k) is crucial for developing a robust model [143].

In the case of spectral data, X will be an m x n matrix, where the m objects are different observations (spectra) and the n variables are the measurements (e.g. wavenumbers) for each object [147]. PCA will separate X into a structured part (S), which basically constitutes the PC-model, and a noise part (E), which measures the lack-of-fit of the model: a smaller E represents a better model [148].

Two more concepts are relevant in PCA: the scores and the loadings. The scores are the values of each object in the new coordinate system, i.e. the projections of each object on the different PCs, and the score matrix T (m x k) constitutes all the scores for all the objects, with the scores of each object making up a row. The loadings, on the other hand, illustrate the weight or importance of each original variable on a given PC. The loading matrix P (n x k) consists of the k PCs retained and acts as the transformation matrix between the original variable coordinate system and the new PC-system. The data matrix X can be decomposed into the scores and loadings matrices [141, 143].

$$X = S + E = TP^T + E \tag{1.1}$$

Generally, the data can be adequately described by far fewer factors than the original variables and hence PCA is highly efficient for qualitative analysis. Intra-sample relationships can be observed by plotting the scores from different PCs against each other and looking for clustering phenomena. In fact, these combinations of variables are often more robust indicators of laboratory sample or process conditions than the original ones individually [140].

Quantitative Analysis: Partial Least Squares Regression

In order to develop a quantitative prediction model, which is the ultimate goal if spectroscopy is to be used to monitor the PHA production process, supervised regression methods are usually applied. In this case, two matrices are necessary, one with the reference information Y (e.g. sample concentration) and the other with the spectra collected, which is basically the previously introduced X matrix. The resulting regression model will be used to predict Y-values from new measurements of X [143].

Partial Least Squares (PLS), for example, is one of the most common tools used with vibrational spectroscopy to estimate and quantify the various components within a sample. PLS is an extension of the PCA method mentioned above, but it carries out its decomposition while looking for a correlation between the X and Y blocks. In other words, the chosen factors, which in PLS are commonly referred to as Latent Variables (LVs), must reduce the dimensionality of the overall data space while being predictive of the information in the dependent block [149]. The result will be a regression matrix that provides the coefficients for the relationship $Y = f(X)$, which can represent, in the end, the relationship between the acquired spectra and the concentration of the analyte of interest.

Model Validation

Lastly, validation is essential. If two many factors are included in the model, it will most likely fit the sampled data perfectly, but will fail to properly predict new data. This phenomenom is called over-fitting. It is, therefore, important that internal validation is first performed to test the robustness of the model. Cross-validation is usually used in this context, leaving out part of the data points, building a model, predicting them back in and doing so repeatedly with different sample sets. Only after ensuring robustness in this manner can one move from the training phase to the validation phase and introduce new measurements for prediction. If the model is proven to have a good prediction strength, it can be used to make future estimations of the property comprised in the Y-block without the need of performing extensive off-line analyses and based exclusively on the spectral data contained in X [143].

2 Aims and Objectives

The present thesis was carried out in the Biochemical Engineering (BIOENG) group, in the Faculty of Sciences and Technology (FCT), Universidade NOVA de Lisboa (UNL), which focuses on developing both fundamental and applied research in the areas of industrial and environmental biotechnology, namely on production and characterization of PHAs.

Furthermore, this work is inserted in the GLOPACK project, which is an on-going Horizon 2020-EU initiative led by the Université de Montpellier and with which the BIOENG group is collaborating. The project focuses on accelerating the transition to a circular economy concept, through the development of innovative packaging solutions with no environmental footprint and the ability to extend the shelf life of food products. The technologies under progress are enabling but simply applicable solutions, such as the development of biodegradable packaging materials issued from agro-food residues.

Hence, the research was integrated in the context of a pilot-scale PHA production plant, using MMCs and fruit pulp waste as feedstock. A 3-stage process was established, which is detailed in the following chapter. In particular, the present thesis aims to follow this pilot-scale production and develop an at-line monitoring strategy for PHA and biomass quantification during the accumulation stage. Both Raman and NIR spectroscopies are to be investigated as potential tools for the development of a quantitative model, using GC as a reference method.

Thus, the experimental approach is based on the collection of a considerable number of samples from different accumulation assays for both spectral acquisition and GC analysis, which will allow for the development of a robust and reliable calibration based on chemometrics tools. Validation with additional samples is also crutial to ensure the validity of the model. Furthermore, a preliminary qualitative analysis is of interest, in order to better understand the data.

Ultimately, if their potential is corroborated by this work, spectroscopy tools may be integrated on-line by means of immersion probes and provide real-time information on the process reactors, thus replacing most of the time-consuming and labor-intensive analytical techniques currently under use. This on-line monitoring will facilitate decision-making during production, namely regarding the number and organic load of the pulses fed during the accumulation.

3 Materials and Methods

3.1 Experimental Set-up

The experimental set-up is illustrated in Figure 3.1. PHA production was carried out through the 3-stage process mentioned in the Introduction (Section 1.4.2), which includes an acidogenic fermentation, followed by the selection and accumulation stages. The set-up and operational parameters of each corresponding reactor are detailed in the following sections.

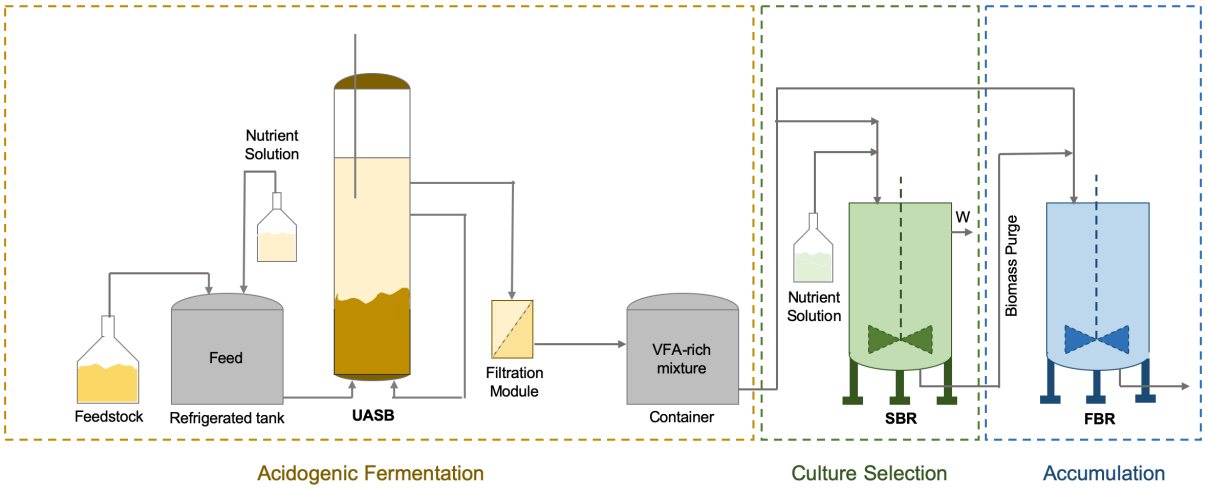


Figure 3.1: Experimental set-up of the 3-stage PHA production process. UASB - Upflow Anaerobic Sludge Blanket; SBR - Selection Batch Reactor; FBR - Fed-Batch Reactor; W - Withdrawal.

It is important to highlight that two different operations of the pilot plant were followed for the present thesis. The first one took place between April and July 2019 (Operation 1) and the second one in October of the same year (Operation 2). All the reactors were shut down during August and restarted in September.

3.1.1 Acidogenic Fermentation

A 100-L acrylic Upflow Anaerobic Sludge Blanket (UASB) reactor was used to carry out the acidogenic fermentation stage, during which apple pulp waste provided by Sumol+Compal S.A. was used to produce a VFA-rich stream. The feedstock was stored at -20°C to prevent degradation and the UASB feed solution was prepared every two or three days in a 500 L refrigerating tank by diluting the apple pulp

with tap water down to the desired substrate concentration. The solution was suplemented with ammonium chloride, NH_4Cl, and potassium dihydrogen phosphate, KH_2PO_4, in order to establish a COD:N:P ratio of 100:0.5:0.1. Furthermore, to maintain granules consistency, cellular structure and activity, calcium as $CaCl_2$, magnesium as $MgSO_4.7H_2O$ and iron as $FeCl_3$ were also added to the medium.

A mixed microbial culture was used to inoculate the reactor. The culture was obtained from an anaerobic digestion process for biogas production from the brewery industry (Super Bock Group, S.A.). Enrichment in acidogenic bacteria was achieved by imposing a low pH-based selective pressure strategy. The reactor was operated continuously with a working volume of 60L and a recirculation flow of 2.4 L/min. The fermentate products outlet was located on the upper part of the reactor and the recirculation and feed streams entered from the bottom and flowed upwards, keeping the biomass granules in suspension. An OLR of 16 g/(L.day) as chemical oxygen demand (COD) was applied to the reactor during Operation 1, while in Operation 2 this loading rate was increased between 10 and 16 g/(L.day) as COD. The final VFA-rich stream was filtered for solids removal before being forwarded to the selection and accumulation reactors.

It should be noted that this reactor was not operated by the author of this thesis and that the description above simply attempts to provide some notion of the pilot-scale process as a whole, since the acidogenic fermentation stage will have significant impact on the subsequent selection and accumulation.

3.1.2 Culture Selection

The culture selection was conducted in a SBR with a working volume of 100 L, operating under a feast and famine regime. The reactor was inoculated with activated sludge from the Mutela WWTP. The operation was carried out in aerobic conditions and the DO levels were monitored through an immersion probe (ISM®, Mettler Toledo). Moreover, a mechanical impeller allowed for stirring and further ensured non-oxygen limiting conditions. There was no control of pH nor temperature, but pH was frequently measured through pH test paper strips and kept around 8 by addition of sodium bicarbonate.

A computer software (BioCTR 59m) was used to control the different phases of the cycle, which was kept at 12 hours. The carbon source was fed in the beginning of the cycle and was mostly composed of the VFA-rich stream provided by the UASB reactor, although synthetic VFA supplementation was often necessary to adjust the OLR to 75 Cmmol/cycle or the ratios of HB and HV precursors and that butyric and valeric acids were used for this purpose. It should be noted that in the present thesis, for supplementation estimations and later polymer characterization, caproate was considered as a precursor of HB monomers. Together with this carbon source, mineral medium with the following composition (mg/L) was added to the reactor: $FeCl_3.6H_2O$: 12, H_3BO_3: 1.2, $CoCl_2.6H_2O$: 1.2, $MnCl_2.4H_2O$: 0.92, $ZnSO_4.7H_2O$: 0.92, $Na_2MoO_4.2H_2O$: 0.46, $CuSO_4.5H_2O$: 0.23, KI: 0.23, ethylenediaminetetraacetic acid (EDTA): 400, $MgSO_4.7H_2O$: 2300 and $CaCl_2.2H_2O$: 300. The nutrient supply was uncoupled from the carbon source and it was performed 2 hours after the beginning of the cycle by feeding ammonium chloride and potassium phosphate in a typical COD:N:P ratio of 100:6.5:1 [Cmmol:Nmmol:Pmmol].

By the end of each cycle, a biomass purge of 12.5 L was removed, the mixed liquor was allowed to settle for 1 hour and 37.5 L of supernatant were discarded. The total volume withdrawn with the purge and after the settling phase was equal to the sum of the volumes added with the carbon, mineral medium and nutrient supplies. Overall, this translated into an SRT of 4 days and an HRT of 1 day.

The SBR performance was monitored through weekly sampling: 6 samples during the feast phase, 1 sample right before nutrient feeding, as well as 5 min and 15 min after that, and 5 more samples cover-

ing the rest of the famine phase. All the samples collected were subjected to GC for PHA quantification and characterization and had their TSS determined for biomass assessment. Furthermore, the samples from the feast phase were analyzed through HPLC for VFA quantification.

Regarding relevant changes applied to the reactor, it should be noted that the SBR was started-up on the 21st of December of 2018 but a washout of the culture ocurred around day 178 and the SBR had to be reinoculated on the 19th of June of 2019. The operation ended on day 43 after this reinoculation. With respect to Operation 2, the SBR was started-up on the 18th of September of 2019 with and OLR of 75 Cmmol/(L.day), which was increased to 150 Cmmol/(L.day) after a week. A synthetic VFA mixture was used during this period, with a HB:HV percursors ratio of 60:40 (%w/w). On day 29, it was observed that considerable levels of ammonia remained at the end of the famine and thus the typical C:N:P ratio was changed to 100:5:1 [Cmmol:Nmmol:Pmmol]. However, after noticing a significant drop in the biomass, the ratio was increased to 100:6:1 [Cmmol:Nmmol:Pmmol] on day 34. The experimental studies were ended on day 40 of this operation.

3.1.3 Accumulation

Once the UASB and SBR operations were stable, accumulation assays were performed. Typically, two purges withdrawn from the SBR after carbon depletion were collected and placed in a 60L fed-batch reactor, which was then fed with the VFA-rich stream in a pulse-wise manner under nutrient limitation. The reactor was aerated and agitated to ensure aerobic conditions and mixing. Similarly to the SBR, there was no automatic control of pH, although the pH was frequently measured through pH test paper strips and kept around 8 by addition of sodium bicarbonate. Furthermore, the reactor was kept at room-temperature.

Each pulse was typically equivalent to the carbon supply of the SBR and was fed to the reactor once the carbon from the previous pulse was entirely consumed. Carbon depletion was indirectly monitored through a DO probe (ISM®, Mettler Toledo), as described in Section 1.4.2 [104]. The accumulation assays were terminated once no significant DO drop was observed after the pulse. Thus, the length of an accumulation and the number of pulses was not fixed, but varied greatly, depending of the performance of each assay.

Monitoring of the accumulation reactor was carried out on a weekly basis, in order to assess the accumulation performance of the selected culture. Hence, samples for GC were collected before and after each feeding pulse to quantify and characterize intracellular PHA, and three extra samples were collected between the first and second and the second and third pulses for a more detailed observation of the accumulation dynamics. Furthermore, the same samples were subjected to HPLC analysis to determine VFA concentration. Lastly, the samples were used to estimate the TSS at each acquisition moment.

Apart from the above mentioned routine monitoring for performance evaluation, samples were collected before each pulse for analysis through Raman and NIR spectroscopy techniques. GC was used as a reference method for PHA concentration (triplicate measurements) and the corresponding TSS were further determined for each data point (duplicate measurements). A total of 16 accumulations were monitored during the first operation of the pilot plant and 13 during the second one. The first set was exclusively analysed through Raman, whereas the second set was investigated through NIR spectroscopy.

3.2 Experimental Procedures

3.2.1 Determination of Suspended Solids

The determination of TSS concentration in the collected samples was carried out according to the standard procedure described in [150]. Samples were filtered through a pre-weighed standard glass-fiber filter, which was then placed in an oven at 105 °C overnight. Once dried, the increase in filter weight represented TSS. The sample volume was generally fixed at 5 mL, although a few samples required smaller volumes of 2, 3 or 4 mL due to higher viscosity. VWR glass-fiber filters with 47 mm in diameter and a particle retention of 1.2 μm were used and the filtration was performed with a vaccuum filtration system. The calculation of this parameter was as follows:

$$TSS = \frac{m_2 - m_1}{V} \tag{3.1}$$

where m_1 refers to the mass of the empty filter, m_2 is the mass after drying at 105 °C and V is the volume of sample subjected to filtration. This method offers an estimation of the amount of matter present in the accumulation reactor, which roughly relates to its biomass concentration.

3.2.2 PHA Quantification and Characterization

GC was used to quantify the polymer produced during the accumulation assays and to characterize it in terms of monomer composition. The protocol for sample preparation was adapted from Serafim *et al* [101], with slight modifications. Before every accumulation pulse, about 4 mL of mixed liquor were collected, centrifuged at 11000 rpm for 3 min, frozen with liquid nitrogen and lyophilized overnight in a Telstar LyoQuest freeze dryer (-51 °C, 0.09 mbar). The resulting pellets were weighed (2-7 mg) in 15 mL test tubes, to which 2 mL of acidic methanol (20% H_2SO_4) and 2 mL of chloroform were subsequently added. The chloroform solution comprised heptadecanoate (HD) at around 1 g/L which acted as an internal standard to improve the accuracy and reproducibility of the method. Tubes were placed in a digestor for 3.5 hours at 100 °C, cooled to room temperature and 2 mL of distilled water were added to each tube, which was then shaken vigorously for 30 seconds by means of a vortex mixer. The two phases were allowed to separate and the organic phase, which comprised the resulting esterified monomers, was extracted and placed in GC vials. Molecular sieves were used for absorption of any residual water.

A gas chromatograph (Varian CP-3800 Gas Chromatograph, Bruker), equipped with a Restek column (60 m, 0,53 mm internal diameter, 1 μm film thickness, Stabilwax) was used for quantification. Helium was used as carrier gas at a flow rate of 1 mL/min and a constant pressure of 14.50 psi. Each run lasted 32 minutes and a volume of 2 μL of sample was injected in the equipment. The temperature started at 40 °C and rose at 20 °C/min up to 100 °C (3 min), followed by a period during which temperature rose up to 155 °C at 3 °C/min (18 min and 20 s), encompassing the elution times of the hydroxyalkanoate monomers. Lastly, temperature was raised up to 220 °C, again at 20 °C/min (3 min and 15 s), and kept for 7 more minutes for cleaning.

Calibration standards were made using an Aldrich copolymer of P(3HB-co-3HV) containing 14% (HV) and 86% (HB) (%(w/w)) with concentrations between 0 and 6.3 g/L. The standards were prepared in an analogous manner to the samples. Squared correlation coefficients (R^2) of 0.987 and 0.994 were achieved for HB and HV calibration curves, respectively. Since HD was used as an internal standard, the resulting standard curves reported the concentration of each monomer as a function of the ratio between

the peak area of the monomer and that of the internal standard. It should be noted that the intercept of the calibration curve was forced to zero. The PHA concentration of the samples was considered as the sum of the concentrations of both monomers, whereas the PHA content in the biomass was calculated by dividing the PHA concentration by the total solids (TS) concentration and, thus, expressed as a percentage. The reason why TSS were chosen as a measure of cell density is that they would give the closest estimation of PHA concentration from this PHA content, which is based on TS.

3.2.3 Volatile Fatty Acids Analysis

VFAs - acetic (Ac), propionic (Prop), butyric (But), valeric (Val) and caproic (Cap) acids - were quantified through HPLC, as described by Wang et al. (2017) [60]. A VWR Hitachi Chromaster equipment with a refractive index detector, a Biorad pre-column (125-0129 30 x 4.6 mm) and Biorad Aminex HPX-87H column (300x7.8 mm) was used for the analyses. Sulphuric acid, H_2SO_4 0.01 N, was applied as eluent at a flow rate of 0.5 mL/min and temperature of 30 °C. A calibration curve with standards ranging between 4-1000 mg/L was used to estimate the concentration of the different organic acids.

3.2.4 Cell Staining and Visualization

Nile blue staining was performed according to Bengtsson *et al.*, 2008 [89] to visualize PHA granules and confirm that the selected culture was effectively enriched with PHA-accumulating organisms. This method was also used to follow an accumulation assay qualitatively by observing the increasing size and number of granules in the biomass. In either case, a drop of Nile blue stain was added to about 1 mL of sample, which was then placed in a thermostatic bath at 55 °C for 15 min. After this incubation period, the sample was examined under a Olympus BX51 epifluorescence microscope (Japan) at 1000X. With this method, intracellular PHA granules could be observed as bright orange fluorescence.

3.2.5 Microbial Classification

In order to obtain a phylogenetic analysis of the mixed bacterial communities in the SBR and accumulation reactor, Fluorescence *in situ* Hybridization (FISH) was carried out as previously described by Amann (1995) *et al.* [151]. Samples were collected from the SBR at the end of the feast phase and immediately fixed with 4% paraformaldehyde (PFA). The fixation protocol was adapted from Nielsen *et al* (2009) [152] and is adequate for Gram-negative bateria. It began with the addition of 1 mL of PFA to 0.5 mL of sample, which was then incubated at 4 °C for 3 hours. After the incubation period, samples were centrifuged at 10000 rpm for 3 min, the supernatant was discarded and the remaining pellet was washed with 1 mL of phosphate-buffered saline (PBS) and centrifuged again. This washing step was repeated once more. Finally, the cells were resuspended in 0.5 mL of absolute ethanol at -20 °C and stored at the same temperature until further analysis.

The next step was to prepare the slides for hybridization. Samples were applied to the wells of specific glass-slides for FISH analysis and placed in an oven at 46 °C for 10 min. Once dried, the samples were dehydrated in a grading series of ethanol solutions at 50%, 80% and 98% (3 minutes in each solution) and dried under compressed air at the end. Probe hybridization was achieved by firstly applying 8 μL of an hybridization buffer (2 M NaCl, 10% sodium dodecyl sulphate (SDS), 1 M Tris-HCl, formamide and miliQ water, pH 8) to each well. The amount of formamide in the buffer was adjusted based on the concentration required by each specific probe and miliQ water was used to make up a

final volume of 2 mL. Then, 0.7 μL of EUBMIX (see below) and 0.7 μL of a specific probe were added to each well. Each slide was placed in a falcon tube containing a moisturising tissue with the remaining hybridization buffer and incubated in the oven at 46 °C for 1.5-3 hours.

The incubation was then followed by a washing step. The washing buffer (2 M NaCl, 10% SDS, 1 M Tris-HCl, 0.5 M EDTA and miliQ water) was prepared in a 50 mL falcon tube and pre-heated to 48 °C in a thermostatic bath. The slides were then place in the buffer tube for 10-15 min, washed with MiliQ water at 4 °C and dried with compressed air. Finally, a few drops of Vectashield mounting media were added to the dried slides - enough to cover all wells. The slides were visualized using ZEISS Axiocam 506 mono fluorescence microscope, at 1000X.

Regarding the probes used, the above mentioned EUBmix is a fluorescein isothiocyanate (FITC)-labelled mixture of probes (EUB338 [153], EUB338II, EUB338III [154]) which together encompass most members of the domain *Bacteria* and are therefore applied to all wells to visualize the entire bacterial community and differentiate it from detritus. Furthermore, specific cyanine 3 (Cy3)-labelled probes for PHA-accumulating bacteria were applied (see Table 3.1). FITC emits in the green wavelength interval, whereas Cy3 does so in the red wavelength region.

Table 3.1: DNA probes used for targeting polyhyroxyalkanoates (PHA) accumulating genera through fluorescent *in situ* hybridization (FISH) analysis.

Probe Name	Targeted Genus	Reference
PAR651	Paracoccus	[155]
Azo644	Azoarcus	[156]
THAU832	Thauera	[157]
LAMP444	Lampropedia	[158]
AMAR839	Amaricoccus	[159]
Zra23a	Zoogloea	[160]
ULB-450	Plasticicumulans	[161]

3.3 At-Line Monitoring

3.3.1 Sampling and Acquisition of Raman Spectra

As mentioned above, different accumulation assays were monitored, with samples collected before every feed pulse. Raman spectra were obtained with a modular spectrometer (Ocean Optics QE65 Pro), coupled with a 785 nm excitation laser (RGBLase LLc, California, USA) that delivered 500mW to the sample. SpectraSuite® was the software platform used for operating the spectrometer. A total of 16 accumulation assays were fully monitored through Raman spectroscopy during the first period of operation of the pilot plant.

A 15 s integration time was applied to all samples. Each spectrum corresponded to a single scan and quadruplicates were measured every time: twice with the same volume of sample and twice more after renovating this volume. Furthermore, a dark spectrum was acquired before each sample and used

as a reference. The spectra X-axis units were defined as Raman shifts, which basically correspond to the change in wavenumber undergone by the monochromatic light scattered while passing through the sample. The detector covered a Raman shift range between -62,34 and 2677,68 cm^{-1} with a 3,69 cm^{-1}/pixel linear dispersion, which in the processed spectrum translated into 1044 data points.

Due to the inherently weak signals obtained in Raman spectroscopy, samples were previously concentrated by centrifugation. Around 40 mL of each sample were collected and centrifuged at 10000 rpm for 3 min. Most of the supernatant was discarded, while ensuring that enough was kept to resuspend the biomass. Graduated pipettes were used both for collecting the sample and discarding the supernatant and thus, the final concentration factor was thoroughly estimated and recorded for later addition to the model. Once homogeneous, the concentrated sample was placed in a 3 mL glass cuvette, with a 1 cm optical path length, and its spectrum was acquired. The pretreament was carried out immediately after collection of the sample from the accumulation reactor and took between 10 to 20 min. The acquisition of the spectra lasted around 10 additional minutes.

3.3.2 Sampling and Acquisition of Near-Infrared Spectra

Just like for Raman monitoring, samples were collected before each feed pulse and analized with a Matrix-F Fourier Transform-NIR Spectrometer (Bruker OPTIK GmbH, Ettlingen, Germany), coupled with a fiber optic probe (Falcata XP 12 NIR, Hellma® Analytics, Müllheim, Germany) with a mechanical pathlength of 1 mm (total optical path length of 2mm). OPUS (Version 8.2) spectroscopy software was used to operate the equipment. NIR spectroscopy was carried out during the second operation period of the pilot plant, on a total of 13 accumulation assays.

Regarding the optical parameters, spectra were acquired in transflectance mode with a resolution of 16 cm^{-1}, a scanner speed of 40kHz and a preamplification gain of B for the samples and A for the background. Quadruplicates were measured for every sample and each spectrum further comprised 64 scans in order to obtain an average and reduce noise. The integration time was dependent on the above mentioned acquisition parameters and was typically around 7 s. Air was used as background and its spectrum was acquired every day before beginning the monitorization (dried empty probe). Overview spectra were acquired in the range of 12,000-4,000 cm^{-1} which, considering a resolution of 16 cm^{-1}, translated into a total of 999 data points per spectra.

No sample pretreatment was performed before spectral acquisition. Around 50 mL of sample were collected from the accumulation reactor and placed in a glass beaker under constant agitation, where the fiber optic probe was directly inserted. The probe was washed with distilled water and ethanol and dried between each measurement. A total NIR spectral measurement time was between 10 and 20 minutes, accounting for sample collection, the sequential acquisition of several spectra and the selection of four consistent ones to be recorded as quadruplicates for every sample.

3.4 Chemometrics

Multivariate data analysis was performed using Matlab version 9.3, R2017b (MathWorks, Natick, MA, USA) and the PLS Toolbox version 8.2.1 (Eigenvector Research Inc., Wenatchee, WA, USA). The Raman data matrix included 95 spectra (rows) and 970 Raman shifts (columns), whereas the NIR data matrix comprised 74 samples (rows) and 999 wavenumbers (columns).

3.4.1 Data Preprocessing

For Raman spectra, the preprocessing stage comprised spectral truncation between 200 and 2000 cm^{-1}, Savitsky-Golay smoothing, followed by SNV normalization and mean-centering. On the other hand, different preprocessing strategies were applied during the optimization of the NIR models, including first and second derivative, SNV normalization and mean-centering and different spectral truncations. A brief explanation of the fundaments of each mathematical preprocessing tool is presented below [140].

Smoothing is done separately on each row (sample) of the data matrix, by assuming that adjacent variables contain redundant information which can be averaged to remove high-frequency noise. The Savitsk-Golay algorithm, in particular, is able to smooth the spectra by successively fitting low-degree polynomials to windows around each data point. In this case, a filter width of 15 points was applied.

The Standard Normal Variate (SNV) normalization is applied to every spectrum individually. It calculates the mean and the standard deviation of all the variables for a given sample and then every data point is subtracted from the mean and divided by the standard deviation. After SNV, each spectrum will have a mean of 0 and a standard deviation of 1. It should be noted that this implies a zero-order detrend (subtraction of the individual mean value from each spectrum), which is, however, different from the mean-centering tool discussed below.

For differentiation, a point-difference first derivative is applied, on which each variable in a sample is subtracted from its immediate adjacent variable. This subtraction leaves only the part of the signal which is different between the two variables, and thus decreases baseline signal and enhances higher-frequency features. The second derivative is calculated by repeating this process. Since differentiation tends to accentuate noise, the above-mentioned Savitsky-Golay algorithm is simultaneously applied as the derivative is taken, with a a filter width of 13 points.

Lastly, mean-centering is included in every preprocessing strategy. This technique calculates the mean of each column and subtracts this from the column. Hence, after mean-centering, each row will include information how that sample deviates from the average sample in the original data matrix.

3.4.2 Principal Component Analysis

PCA was performed on both Raman and NIR data matrices for data compression and information extraction. Mathematically, PCA uses an eigenvector decomposition of the covariance of the process variables. For a given mean-centered data matrix X with m rows and n columns, the covariance matrix of X is defined as:

$$cov(X) = \frac{X^T X}{m-1} \tag{3.2}$$

The eigenvectors of this covariance matrix, p_i, correspond to the loadings in the following decomposition of matrix X:

$$X = t_1 p_1^T + t_2 p_2^T + t_3 p_1^T + ... + t_k p_k^T + E \tag{3.3}$$

where k is the chosen number of factors or PCs, E is the residual matrix and t_i are the scores vectors, which are basically the projection of X onto the p_i. The t_i, p_i pairs are arranged in descending order according to the corresponding eigenvalue, λ_i, which is a measure of the amount of variance captured by the pair.

For outlier detection, a 95% confidence limit was established on the scores plots, which can be calculated by assuming a normal distribution of the scores and applying the student's t-distribution. Furthermore, the Q lack-of-fit statistic was used to evaluate how each sample conformed to the PCA model: for a given sample, Q represents the square of the corresponding row of matrix E. Lastly, the Hotelling's T^2 statistic was employed as a measure of the variation in each sample within the PCA model.

A more detailed explanation of the statistics associated with PCA models can be found in the PLS Toolbox Chemometrics Tutorial [140].

3.4.3 Partial Least Squares Regression

Of the 74 NIR samples, 4 were excluded as outliers and 57 were used in PLS regression for quantitative determination of PHA content (%), PHA concentration (g/L) and TSS concentration (g/L). The remaining 12 samples were used for test set validation, after performing a full cross-validation and preselecting the calibration model. The PLS toolbox default cross-validation method was applied: venetian blinds with 8 splits and 1 sample per split. Calibration models were evaluated based on the squared coefficient of determination (R^2), root mean square errors (RMSE) of calibration, and cross-validation (RMSEC and RMSECV, respectively). The optimal number of LVs was identified as that leading to the minimal RMSECV and a total of 20 LVs were estimated for each calibration. The final models were chosen based on their root mean square error of prediction (RMSEP). The mathematical expression for the RMSE is presented below (Equations 3.4).

$$RMSE = \sqrt{\frac{1}{m}.\sum_{i=1}^{m}(Y_i^{meas} - Y_i^{pred})^2} \tag{3.4}$$

where m is the number of samples, Y_i^{meas} are the reference values and Y_i^{pred} are the values predicted in every situation, calibration, cross-validation or test set prediction.

Regarding the mathematics of the method, PLS attempts to find LVs which maximize the amount of variation explained in X (e.g. spectra) that is relevant for predicting Y (e.g. reference values). The SIMPLS algorithm was applied for the present work, which focuses on maximizing the covariance of these two matrices, also through an eigenvector decomposition. For more on SIMPLS-related algorithms, see de Jong, 2001 [162]. Finally, the same statistical tools were employed to establish confidence limits within the model.

4 Results and Discussion

4.1 Overview of Process Performance

The three-stage process described in Chapter 3 was successfully applied to the pilot-plant in two operational periods. The set-up allowed for culture enrichment in PHA-accumulating organisms and ultimately enabled the pilot-scale production of a P(3HB-co-HV) copolymer. The focus of this work is on the accumulation stage, although punctual references to the SBR will be made during this chapter in order to explain some of the effects observed in the accumulation assays. Tables 4.1 and 4.2 summarize the most relevant parameters of every accumulation monitored during Operation 1 and 2, respectively.

4.1.1 Production Outcome

Regarding Operation 1, it is worth highlighting that the SBR had to be started-up again after a wash-out of the culture took place. Since the UASB reactor was stable at the time, all the accumulations used its VFA-rich outlet stream as feed, although supplementation was often carried out in order to increase its organic load or adjust the HB:HV ratio. Only two accumulations were monitored before the wash-out and therefore the bulk of the results of this operation period rely on the assays performed after reinoculating the SBR. Overall, the percentage of HV in the copolymer varied between 17 and 26 % (% w/w) and PHA contents up to 78 % (%w/w) were achieved. High cell densities for MMC standards were also obtained, up to 22 g/L as TSS. The results in terms of PHA storage are comparable to the highest storage capacities in the literature and the cell densities are considerably higher than those reported so far (see Table 1.1).

Operation 2, on the other hand, overlapped with the start-up of the entire 3-stage process and, therefore, the UASB reactor was not yet stable when the accumulation assays began. Hence, the first assays of this operation period used a synthetic VFA mixture as feed, which was changed to the UASB fermentate outlet as soon as possible. Supplementation was necessary in most of the assays, mainly to increase the VFA load but also to adjust the HB:HV ratios, as mentioned above. However, the last two accumulations used the VFA-rich outlet exclusively as feed. It is highlighted that the SBR was fed with a synthetic VFA mixture during the entire monitorization period, regardless of the type of feed used for the accumulation assays. The HV content in the copolymer produced was slightly higher than that of Operation 1, with a range between 22 and 35 %. The two last accumulations were an exception (see Accumulation 12b and 13b in Table 4.2) since the HV content went up to 52 %. This was probably the result of not supplementing the fermentate feed, which at the time presented a higher fraction of HV precursors than expected. Nevertheless, similar PHA contents were obtained during this second operation, with a maximum of 89 % (%w/w) reported - or 71 % if these last couple of accumulation assays were to be excluded. However, lower cell densities, of up to 11 g/L, were observed.

Table 4.1: Summary of the most important parameters regarding the accumulation assays carried out during Operation 1. The SBR was first started-up on the 21st of December, 2018 (accumulations labelled with roman numerals) and then again on the 19th of June, 2019 (accumulation labelled with arabic numerals). The VFA composition in terms of acetic (Ac), propionic (Prop), butyric (But), valeric (Val) and caproic (Cap) acids was estimated in a carbon-molar basis (in Cmmol/L) and took supplementation into consideration. In some cases, (1) two or (2) three purges of the SBR were used in the accumulation assay. n.a. stands for not available.

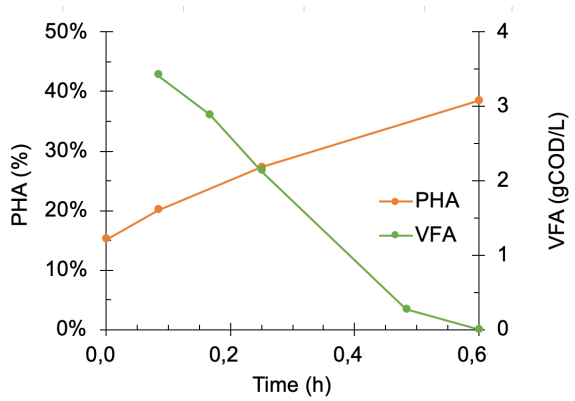
Accumulation Label	Day since SBR Start-up	Feed Type	VFA Load (Cmmol/L)	VFA Composition (Ac:Prop:But:Val:Cap)	Number of Pulses	Total Feed Volume (L)	PHA Content (%, gPHA/gTS)	HB:HV Ratio (%w/w)	TSS (g/L)
i	167	Real	260	32:5:46:6:11	5	40	72	76:24	4.43
ii	172	Real	n.a	n.a.	8	46	75	79:21	9.34
1a	13	Real	n.a	n.a.	7	32	n.a.	76:24	7.77
2a	14	Supplemented	n.a	n.a.	5	16	n.a.	76:24	9.80
3a	16	Supplemented	850	18:6:66:10:0	7	15	77	78:22	12.3
4a	21	Supplemented	560	12:4:73:11:0	7	25	45	77:23	16.8
5a	22	Supplemented	860	14:4:71:11:0	6	20	53	76:24	11.7
6a	27	Real	250	31:6:54:9:0	6	45	64	68:32	15.1
7a	28	Real	250	31:6:54:9:0	6	36	42	74:26	14.1
8a	29	Real	250	31:6:54:9:0	6	40	44	78:22	16.1
9a(1)	34	Real	250	31:6:54:9:0	4	60	n.a.	79:21	12.0
10a	35	Real	250	31:6:54:9:0	8	34	53	83:17	18.4
11a	36	Supplemented	1010	6:1:80:13:0	8	14	56	80:20	15.8
12a	37	Supplemented	1010	6:1:80:13:0	6	15	42	83:17	12.0
13a	40	Supplemented	1010	6:1:80:13:0	8	19	78	79:21	16.9
14a(2)	43	Supplemented	820	7:2:78:13:0	5	45	69	80:20	22.4
15a	43	Supplemented	820	7:2:78:13:0	7	21	72	76:24	14.3

Table 4.2: Summary of the most important parameters regarding the accumulation assays carried out during Operation 2. The SBR was started-up on 18th of September, 2019. The VFA composition in terms of acetic (Ac), propionic (Prop), butyric (But), valeric (Val) and caproic (Cap) acids was estimated in a carbon-molar basis (in Cmmol/L) and took supplementation into consideration. Accumulation 9b used a purge collected after the SBR feast phase, which is equivalent starting the accumulation after a first feeding pulse.

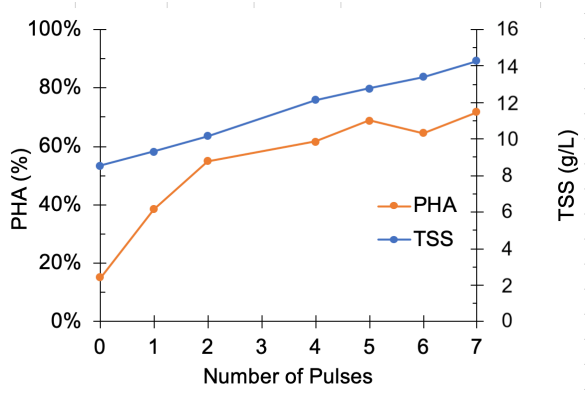
Accumulation	Day since SBR Start-up	Feed Type	VFA Load (Cmmol/L)	VFA composition (Ac:Prop:But:Val:Cap)	Number of Pulses	Total Feed Volume (L)	PHA Content (%, gPHA/gTS)	HB:HV Ratio (%w/w)	TSS (g/L)
1b	19	Synthetic	580	55:11:20:14:0	6	13	50	67:33	6.24
2b	20	Synthetic	580	55:11:20:14:0	5	10	51	65:35	6.39
3b	21	Supplemented	675	3:2:77:18:0	5	10	47	76:26	8.31
4b	23	Supplemented	675	3:2:77:18:0	5	16	55	77:23	11.1
5b	27	Supplemented	675	3:2:77:18:0	5	13	52	78:22	9.69
6b	28	Supplemented	675	4:3:75:17:1	4	8	56	74:26	6.35
7b	29	Supplemented	675	4:3:75:17:1	5	8	60	73:27	7.19
8b	30	Supplemented	675	4:3:75:17:1	5	13	62	74:26	8.79
9b	33	Supplemented	675	2:1:76:19:2	4	4	71	77:23	3.58
10b	34	Supplemented	675	2:1:76:19:2	5	6	63	77:23	3.34
11b	35	Supplemented	675	2:1:76:19:2	5	8	65	77:23	6.29
12b	37	Real	160	5:3:37:36:19	7	28	75	48:52	3.63
13b	40	Real	160	5:3:37:36:19	9	43	89	49:51	4.57

It is worth highlighting the signifcant drop in TSS levels observed on day 33, which was certainly due to the changes in C:N:P ratio applied to the SBR (100:6.5:1 to 100:5:1) on day 29. It appears that there was some serious nitrogen limitation after a few cycles, which eventually resulted in a significant loss of biomass. Although the ratio was increased right after this observation (day 34), the culture was not able to go back to the initial TSS concentrations during the monitored operation period.

Regarding the dynamics of an accumulation assay, Figure 4.1 below illustrates the typical trends in terms of PHA production, TSS growth and VFA consumption, and uses accumulation 15a as an example. It can be noticed that the PHA content increased with VFA consumption during the first pulse and that VFAs were entirely consumed by the end of the pulse (Figure 4.1 (a)). Regarding the accumulation as a whole, the PHA production rate was higher during the first few pulses, generally approaching a plateau by the end of the assay (Figure 4.1 (b)). As was expected, PHA accumulation was accompanied by a considerable increase in TSS concentration.



(a) PHA content (%) and VFA concentration (gCOD/L) versus time during the first accumulation pulse.



(b) PHA content (%) and TSS concentration (g/L) versus number of pulses in an accumulation assay.

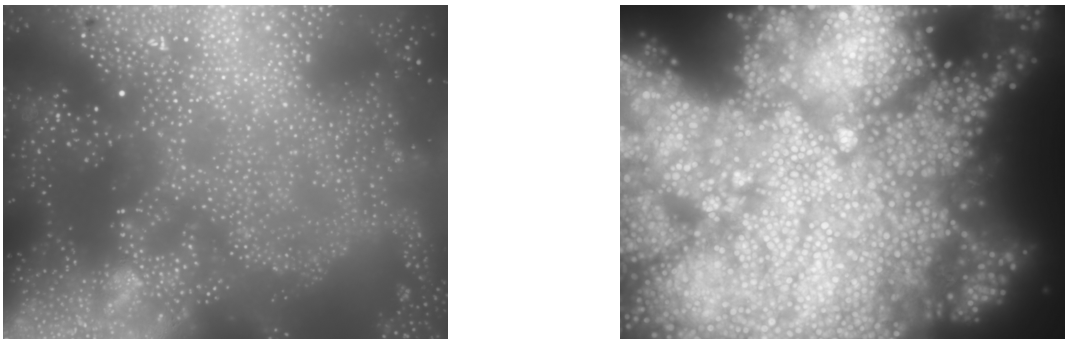
Figure 4.1: Typical trends (a) during the first accumulation pulse in terms of PHA content (%) and VFA concentration (gCOD/L) versus time elapsed since the pulse feeding; (b) throughout the accumulation regarding PHA content (%) and TSS (g/L). Samples in (b) were collected before every pulse feeding. The data corresponds to an accumulation carried out on day 43 of Operation 2 using the same feed and purge as that of Accumulation 14a.

It is worth mentioning, however, that a dilution effect was often detected, mostly when no supplementation was carried out and, thus, a bigger volume was fed to the reactor. This dilution sometimes cancelled the increase in PHA and TSS concentration. Furthermore, a considerable number of accumulation assays in Operation 2 revealed a slight drop in PHA content with the last pulse, insinuating that the PHA-accumulating organisms might be reaching their storage limit and, eventually, starting to consume the polymer. A detailed summary of the PHA content and TSS concentration during each accumulation assay in Operation 2 can be found in Appendix A.

4.1.2 Mixed Microbial Culture Dynamics

In order to ensure the presence of PHA-accumulating microrganisms in the culture and to confirm the production of intracellular PHA granules, Nile Blue staining was performed on samples collected during an accumulation assay. Figure 4.2 illustrates the difference in polymer storage between the beginning (Figure 4.2a) and the end (Figure 4.2b) of the accumulation stage. The fact that fluorescence is considerably stronger in the latter indicates that PHA was, indeed, being stored intracellularly, since Nile

Blue is known to be specific for PHA granules and the granules emit strong fluorescence when they are bound to this fluorescent dye.



(a) Before the accumulation assay. **(b)** After the accumulation assay

Figure 4.2: Fluorescence images of the biomass samples taken from the accumulation reactor on day 30 (Accumulation 8b), at 1000X. Image a) refers to the sample taken before the first pulse and (b) corresponds to the final accumulation point, taken by the end of the fifth pulse. The images show the PHA granules stained with Nile Blue, which are enhanced under fluorescent light.

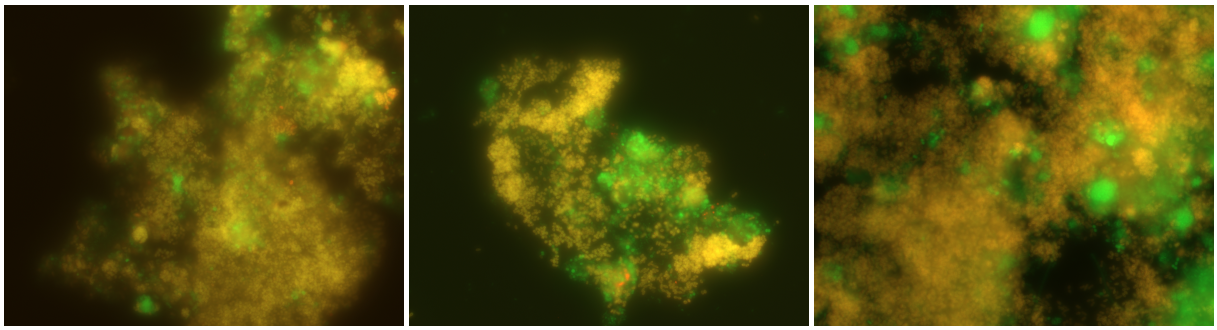
Furthermore, an analysis of the dynamics of the culture was carried out in order to infer if it had eventually reached a steady state in terms of its bacterial community. Regarding Operation 1, this analysis relied exclusively on the SBR performance parameters, namely on the F/F ratio and on the PHA production and VFA consumption rates. These parameters were obtained through weekly monitorings of the reactor and are summarized in Table 4.3. The significant decrease in the F/F ratio, along with the increase in both PHA production and VFA consumption rates over time indicate that the culture was under enrichment in PHA-accumulating microrganisms. Overall, it can be concluded that the bacterial community was dynamically changing during most of the first operation period, since data regarding the last week of operation is not available.

Table 4.3: SBR performance parameters during Operation 1: feast/famine ratio, PHA production rate and VFA consumption rate. The PHA production rate is a global rate, which takes into account the difference in intracellular PHA between the beginning and the end of the feast phase and divides it by the length of this phase, whereas the VFA consumption rate was determined based on the slope of the VFA consumption curve during the same period.

Monitoring	Day since SBR Start-up	Feast/Famine Ratio	PHA Production Rate ($gCOD.L^{-1}.h^{-1}$)	VFA Consumption Rate ($gCOD.L^{-1}.h^{-1}$)
1a	8	0.268	1.50	0.70
2a	16	0.064	2.74	5.18
3a	28	0.050	4.11	7.40
4a	35	0.037	7.78	9.30

For the second operation period, on the other hand, FISH was performed on a biomass sample collected from the SBR on day 6. The culture was screened for PHA-producing bacteria and it was possible to conclude that the selection process was effective in only a few days. In fact, the bacterial community was dominated by the genus *Azoarcus*, with a strong presence of the genera *Thauera* and *Plasticiumulans* and some traces of *Amaricoccus* and *Paracoccus*. Figure 4.3 shows some of the corresponding FISH images. With the (FTIC)-labelled EUBmix probes emitting in the green wavelength region

and the specific (Cy3)-labelled probes emitting in the red one, the targeted PHA-accumulating bacteria are depicted in orange. No further FISH analyses were carried out after this day and therefore a dynamic study of the culture was not possible.



(a) Cy3-labelled Azo644 probe **(b)** Cy3-labelled THAU843 probe **(c)** Cy3-labelled ULB450 probe

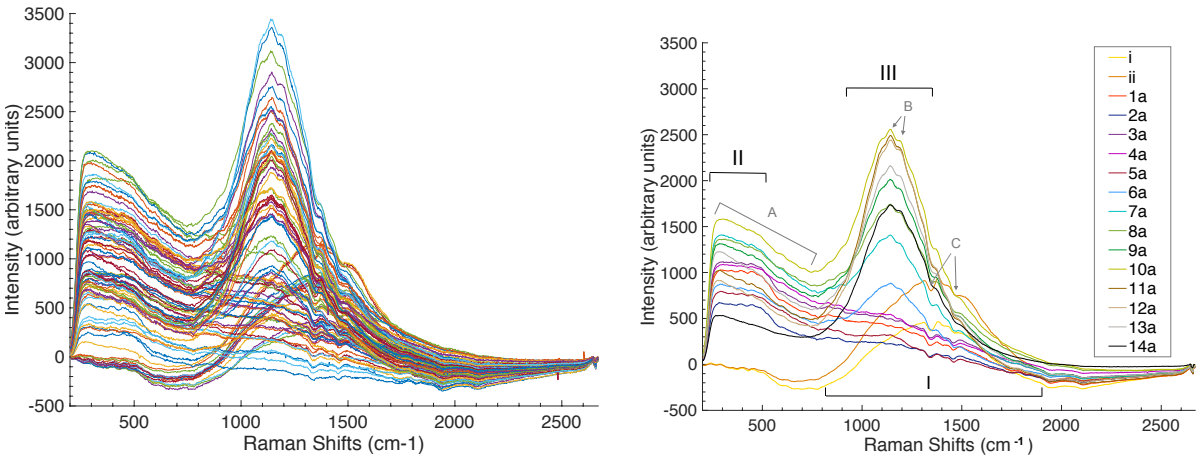
Figure 4.3: FISH images of the biomass samples taken on day 6 of operation of the SBR, at 1000X. The images show bacteria hybridized with FITC EUBmix probe (green) and Cy3-labelled probes (red): Azo644 (a); THAU832 (b) and ULB450 (c). An orange coloration is observed where the general and specific probes overlap, corresponding to the targeted cells.

4.2 Process Monitoring through Raman Spectroscopy

4.2.1 Preliminary Spectral Analysis

To study the possibility of using Raman spectroscopy as a monitoring tool in the production process, samples were collected before each pulse in several accumulations and pretreated as described in Subsection 3.3.1. It was assumed that the VFAs present in the feed were entirely consumed by the time of each sample collection and, therefore, samples were exclusively analysed in terms of TSS concentration and PHA content (see procedure in Section 3.2). Given the inherently weak signals which are characteristic of this technique, the pretreatment mostly relied on concentrating the samples before acquisition. A total of 95 spectra were considered for this study, all of which are depicted in Figure 4.4 (a). Each spectrum was taken as the average of quadruplicate measurements and a total of 16 accumulations were comprehended in this data set.

The first thing to notice is that the shape of the spectra is not constant among the different samples, although it appears to be coherent within a given accumulation and even within a given week. In order to establish some trend in this variation, the mean spectrum of every accumulation was estimated and is represented in Figure 4.4 (b). Firstly, there is a clear distinction between accumulations i and ii, hereby Group 1, and those who took place after reinoculating the SBR, as the former exhibit a wide band around 1400 cm^{-1} which is non-existent in any of the others (I). Regarding accumulations 1a to 14a, there also seems to be an obvious division: accumulations 1a to 5a (Group 2) show a plateau within the first 500 cm^{-1} (II) and then a descending trend along the rest of the spectral range, whereas accumulations 6a to 14a (Group 3) display the same initial elevation but also a clear peak at around 1200 cm^{-1} (III). It actually appears that the signal is evolving in terms of this latter peak, which is absent in the first accumulations and then starts to increase over time, eventually becoming predominant in the spectra.



(a) Raw spectra of the 95 samples in the data set.

(b) Mean spectra of every accumulation in the data set.

Figure 4.4: Raw Raman spectra obtained from at-line monitoring of the accumulation bioreactor during Operation 1. (a) refers to all 95 spectra in the data set, while (b) displays the mean spectra of each of the 16 accumulations. I,II and III indicate the most evident spectral bands, while A, B and C point out some subtle peaks on top of these major bands which seem transversal to the different accumulations. The region below 200 cm^{-1} was discarded.

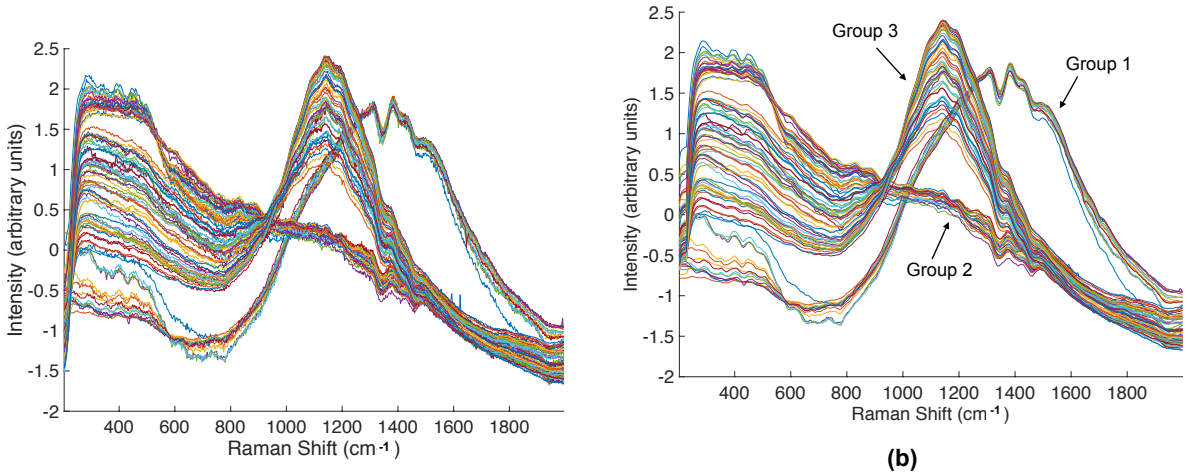
Since there was no relevant change in the operational parameters of the process (e.g. feeding strategy), nor any proportional trend in the accumulation performance (e.g. PHA content or TSS concentration) during this period, one possible explanation for this signal variation might be in the culture's dynamics. As was briefly mentioned before, the results gathered in Table 4.3 clearly indicate that the culture in the SBR was evolving during Operation 1 and thus, there might have been significant changes in the bacterial community to justify the different spectral shapes detected. In fact, it has been reported in the literature that Raman spectra can be treated as fingerprints, providing information on the structure and composition of microbial cells [163]. Unfortunately, as mentioned above, no FISH data is available on this operation period to validate this hypothesis.

On the other hand, it is interesting to see that despite these broad bands that dominate the spectra, a few subtle peaks can be identified on top of these bands which are transversal to the different accumulations. By inspecting Figure 4.4 (b) once more, one can see that the first plateau region presents the same profile throughout the accumulations, up to a Raman Shift of around 800 cm^{-1} (A). Furthermore, two subtle peaks between 1100 and 1200 cm^{-1} can be detected both on top of the major band in Group 3 as on the other groups which lack this band (B). Similarly, two small depressions can be identified between 1300 and 1500 cm^{-1}, which stand out on top of the characteristic broad band of Group 1 but are also subtly present in the others (C). This might indicate that the Raman spectra might indeed be capturing some common structures beyond these most evident global changes.

4.2.2 Preprocessing Strategy

The spectral range was limited from 200 to 2000 cm^{-1} because the regions outside this interval showed excessive noise. Nevertheless, and despite the averaging of quadruplicate measurements, a significant degree of noise can still be detected. Furthermore, the intensity of the signal was highly variable during acquistion, apparently indicating very low reprodutibility among replicate samples or even consecutive measurements of the sample. This intensity changes might have been caused by laser power fluctuations, since the equipment was believed to be close to its end-of-life. No fluorescence

was detected, however, which was expectable since a red (785 nm) excitation laser was used and fluorescence is common with higher frequency lasers. Given these considerations, a simple preprocessing strategy based on smoothing followed by standard normal variate (SNV) normalization was applied to the raw spectra, to reduce random noise and compensate for intensity shifts, respectively. Figure 4.5 shows the visual effects of these preprocessing tools on the spectra. All spectra were further mean-centered before moving on to the subsequent qualitative analysis. The ultimate goal of the preprocessing stage is to remove variation within the data that does not relate to the analytical information of interest.



(a) SNV normalization

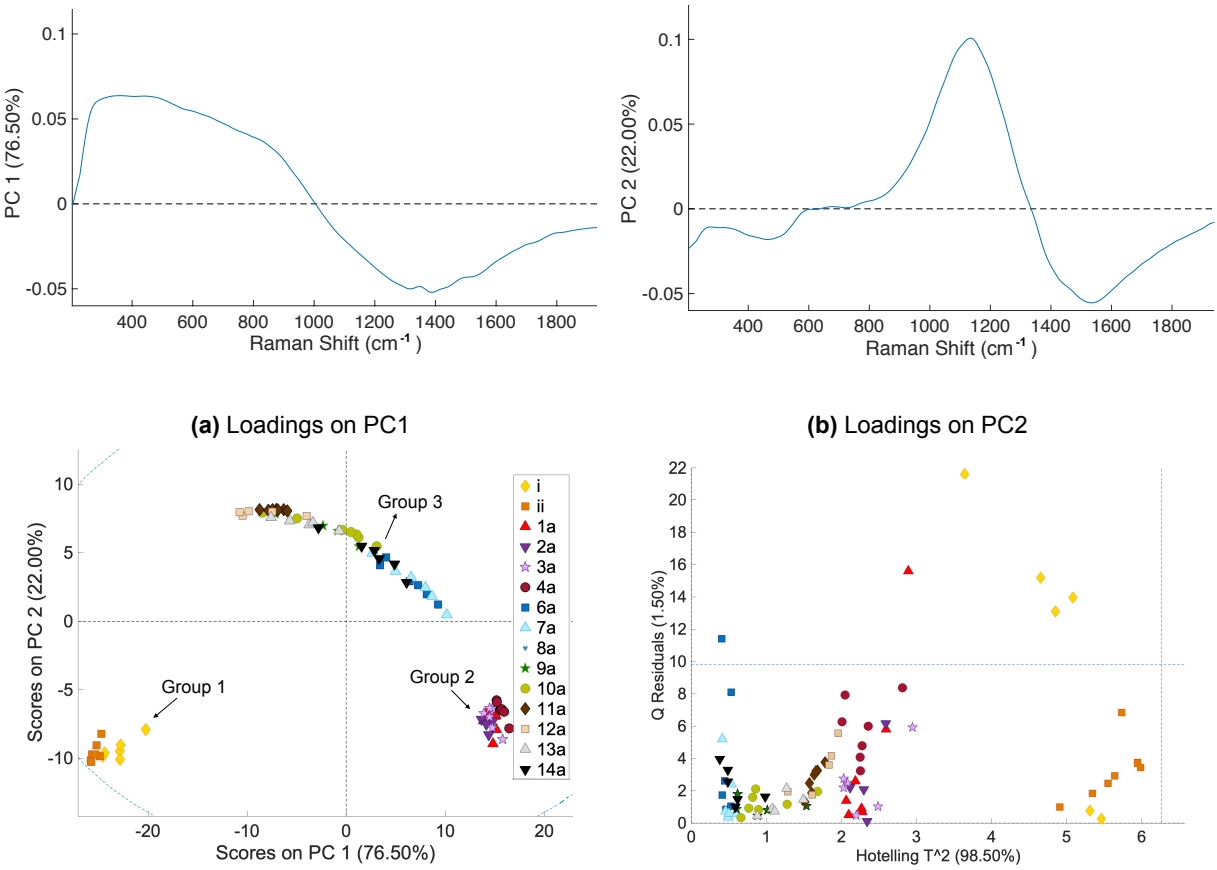
(b) Savitsky-Golay smoothing + SNV normalization

Figure 4.5: Preprocessed Raman spectra obtained from at-line monitoring of the accumulation bioreactor during Operation 1. (a) refers to using SNV normalization alone and (b) corresponds to a Savitsky-Golay smoothing step followed by SNV normalization. Three different groups are identified on (b), based on their spectral shape: Group 1 includes accumulations i and ii, Group 2 refers to accumulations 1a to 5a and Group 3 comprehends accumulations 6a to 14a. Spectra are truncated between 200 and 2000 cm^{-1}.

4.2.3 Principal Component Analysis

In order to extract further information from the spectral data, a PCA was performed. The 970 Raman Shifts detected by the spectrometer were inserted as variables, along with the 95 spectra as samples. Through PCA, this high number of variables was successfully reduced to a few factors, which are combinations of the original variables and are capable of describing major trends in the data. In fact, two PCs were enough to capture 98.5 % of the variance in the data. In this decomposition process, three important data sets arose: the loadings, which describe how each PC relates to the original variables (Figure 4.6 (a) and (b)), the scores, which represent each sample in the new coordinate system (Figure 4.6 (c)) and the residuals, which include the variance that was not captured in the model (Figure 4.6 (d)).

The scores plot (Figure 4.6 (c)) clearly points out the grouping that was predicted above, with Group 1 scoring very low both on PC1 and PC2, Group 2 with high scores on PC1 and low scores on PC2 and Group 3 scoring high on PC2 and spreading across PC1. This was expected, since PCA looks for the greatest amount of variation in the data, which in this case would undoubtedly be the changing shape of the spectra. In fact, this can be sustained by comparing the loadings plots and the preprocessed spectra (Figure 4.5 (b)). Figure 4.6 (a), indicates that PC1 takes into account the initial plateau region, explaining the high scores for Group 2 and the dispersion on Group 3. Furthermore, the band around 1400 cm^{-1}, or in this case the absence of it, seems to have a considerable weight on this PC, explaining the low scores for Group 1. The loadings of the second PC, on the other hand, indicate that this variable



(a) Loadings on PC1

(b) Loadings on PC2

(c) Scores plot: PC2 against PC1

(d) Q residuals against Hotelling T^2 statistics.

Figure 4.6: PCA model on Raman spectra obtained from at-line monitoring of the accumulation bioreactor during Operation 1. The preprocessing strategy consisted on Savistky-Golay smoothing followed by SNV normalization and mean-centering. The spectral range was constrained between 200 and 2000 cm^{-1}. The dashed blue lines in (c) and (d) represent 95% confidence limits.

is mostly influenced by the peak between 1100 and 1200 cm^{-1}, which provides Group 1 and 2 with low scores and causes further segregation among Group 3.

Figure 4.6 (d) is particularly useful in outlier detection, as it provides a 95 % confidence limit. Essentially, there are two types of outliers, those that result in an exceptionally large score and are, therefore, far from the rest of the samples in the scores plot, and those with a large residual because they do not match the PCA model well. A plot of Q residuals vs. Hotelling's T^2 statistics gives a compact view of both residual and score outliers, respectively. In this case, samples belonging to Group 1 have the highest Hotelling's T^2 values, some of which are simultaneously outside the Q residuals confidence interval. It is possible to conclude that the model is particularly inadequate for this group of samples, possibly because the number of data points belonging to this group (14 out of 95) is too low to provide the model with the necessary robustness.

A more detailed inspection of the scores on PC1 can be carried out on Figure 4.7. Although no distinction whatsoever is made within Group 1 and 2, it seems like there is a clear trend among the samples of Group 3, with increasing PC1 scores along each accumulation. This tendency might be due to a number of factors, from increasing cell density or PHA content to the build-up of feed particles, or even some subtle changes in the overall structure and composition of the biomass. Nonetheless, this

shows significant potential that ought to be further explored through a quantitative analysis, such as PLS regression.

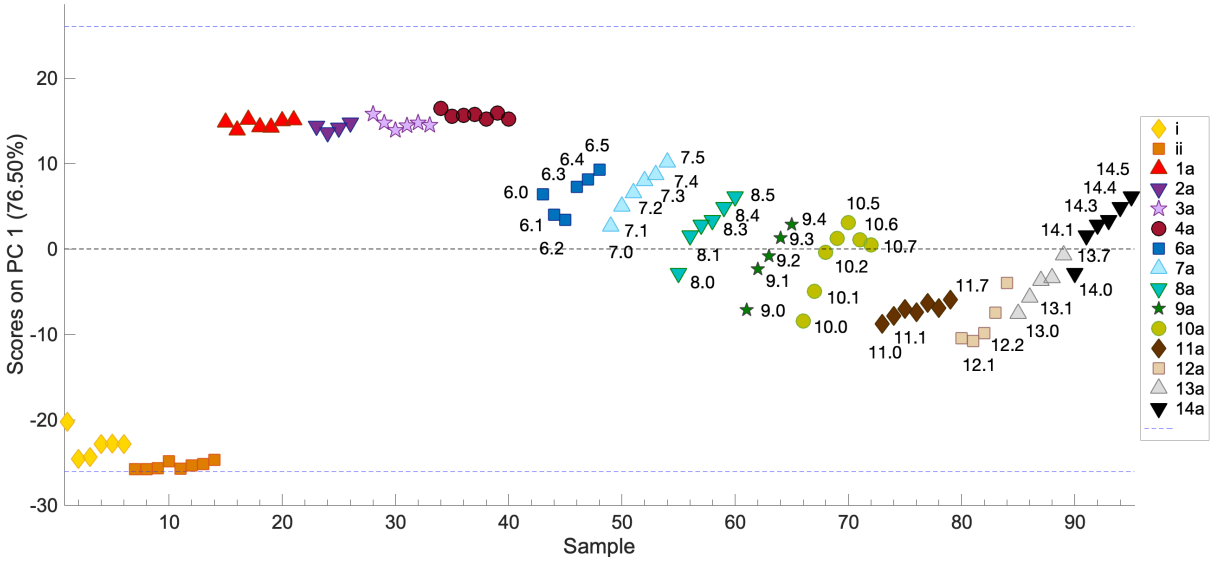


Figure 4.7: PC1 scores from the PCA model on Raman spectra obtained from at-line monitoring of the accumulation bioreactor during Operation 1: PC1 against sample number. The dashed blue line represents a 95% confidence limit.

Furthermore, the identification of the above-mentioned subtle peaks that are transversal to the different sample groups reinforces that, provided with a considerable number of samples from each group, a quantitative model is worth pursuing. In fact, with the right spectral truncation, the grouping effect might simply be a surmountable complication. Unfortunately, the reference values in terms of PHA content obtained throughout Operation 1 are not accurate and, therefore, the development of a PLS model based on them would most likely lead to unreliable results and conclusions.

4.3 Process Monitoring through NIR Spectroscopy

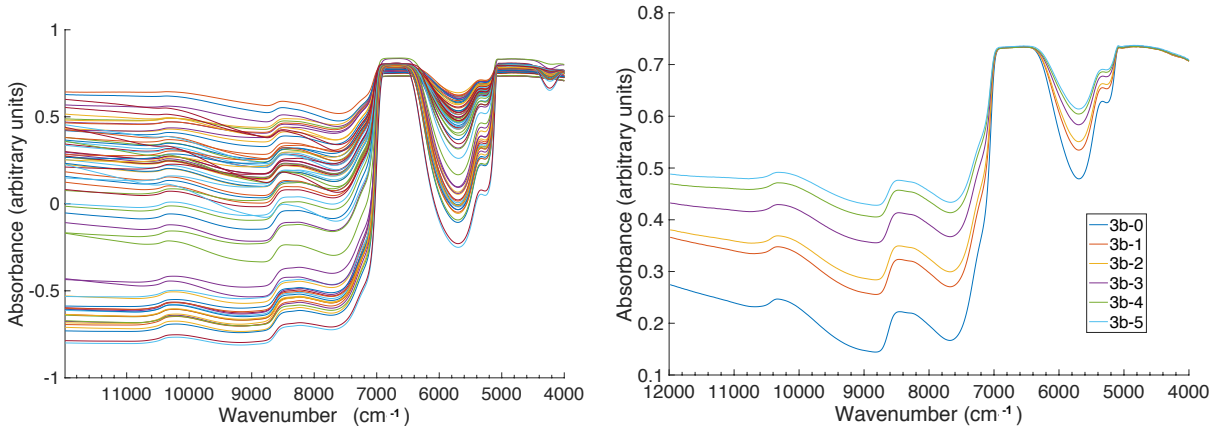
4.3.1 Preliminary Spectral Analysis

As in the previous section, to study the possibility of using NIR spectroscopy to monitor the production process, samples were collected before each accumulation pulse. In this case, no pretreatment was required and the collected samples were directly measured with the NIR spectrometer, since the NIR signal is significantly stronger than that induced by Raman scattering. A total of 74 spectra were used as input, all of which are displayed in Figure 4.8 (a). Each of the represented spectra consists of the average of quadruplicate measurements and a total of 13 accumulations were included in the analysis.

NIR spectra of microbial cultivation broths are generally challenging to analyze given their complex matrices. In this case, further complexity is added, since the analyte of interest is intracellular. By inspecting Figure 4.8 (a), two saturation regions can be identified, between 7000-6400 cm^{-1} and 5100-4600 cm^{-1}, which correspond to the O-H bond first overtone and combinations, respectively. In fact, it is often the case with aqueous matrices that most of the relevant regions are overshadowed by water absorption bands. Furthermore, the increasing baseline shift that can be observed above 8000 cm^{-1}

is probably the result of light scattering by the particles in suspension, namely biomass [164]. In fact, suspended solids are known to induce light diffusion, decreasing the intensity of the incident radiation reaching the detector and thus raising absorption [165]. Figure 4.8 (b) shows the typical trend observed in the absorption spectra throughout an accumulation assay. The baseline shift becomes very clear, reinforcing the hypothesis that it results from increasing cell density.

Generally, raw NIR spectra do not provide information in a clear and direct manner, since they are the result of overlapping absorption bands and the above mentioned light scattering differences. They must, therefore, go through an appropriate preprocessing strategy before being used to construct calibration models, so as to enhance their spectral features and enable the extraction of relevant chemical and physical information.



(a) Raw NIR spectra of the 74 samples in the data set.

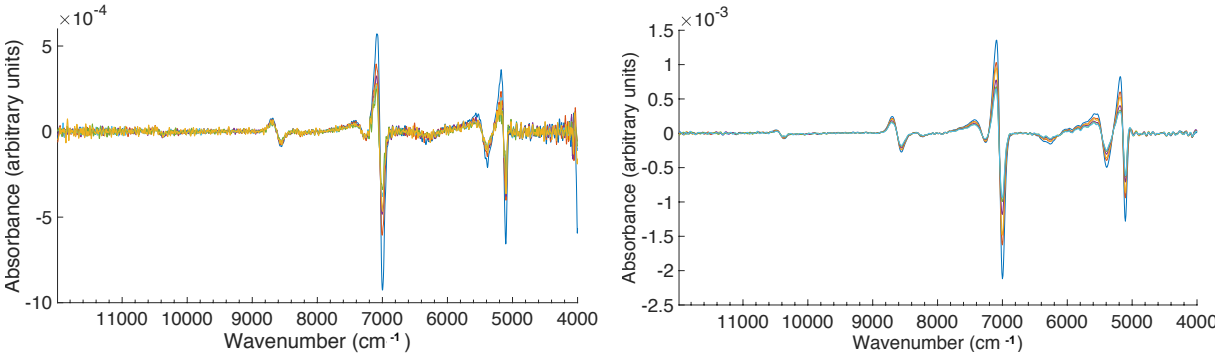
(b) NIR spectra of the samples from accumulation 3b.

Figure 4.8: Raw NIR spectra obtained from at-line monitoring of the accumulation bioreactor during Operation 2. (a) refers to all 74 spectra in the data set, while (b) depicts the spectra from all the samples in accumulation 3b to illustrate the typical trend along an accumulation.

4.3.2 Spectra Acquisition and Preprocessing

As a primary step, only the second derivative spectra were calculated, since this represents one of the most common preprocessing techniques found in the literature for NIR spectra of biomass samples and, in particular, of that containing intracellular PHA [138, 122]. In fact, the second derivative is especially effective in deconvoluting broad overlapping peaks and removing baseline trends.

It was by inspection of these second derivative spectra that it was possible to optimize the acquisition parameters. Figure 4.9 (a) displays the second derivative of spectra corresponding to a preliminary test accumulation. These were acquired with a resolution of 8 cm^{-1} and a scanner velocity of 10kHz, which clearly led to a severe degree of noise. In contrast, Figure 4.9 (b), includes spectra obtained with new and optimized acquisition parameters: noise was decreased at the cost of resolution, which was changed to 16 cm^{-1}, whereas the scanner velocity was increased to 40kHz, to compensate for some interferences arising from sample agitation.



(a) Resolution of 8 cm^{-1} and scanner velocity of 10kHz. **(b)** Resolution of 16 cm^{-1} and scanner velocity of 40kHz.

Figure 4.9: Second derivative NIR spectra from at-line monitoring of the accumulation bioreactor during Operation 2, obtained with different acquisition parameters.

The second-order derivative is also characterized by a negative band with minimum at the same wavelength as the maximum on the original spectrum. For this reason, the differentiated spectra were used to carry out a detailed analysis of the NIR signal and to identify the various functional groups responsible for each of the absorption bands. Figure 4.10 portraits the second derivative spectra, on which each spectral region is properly identified with the corresponding overtone or combination.

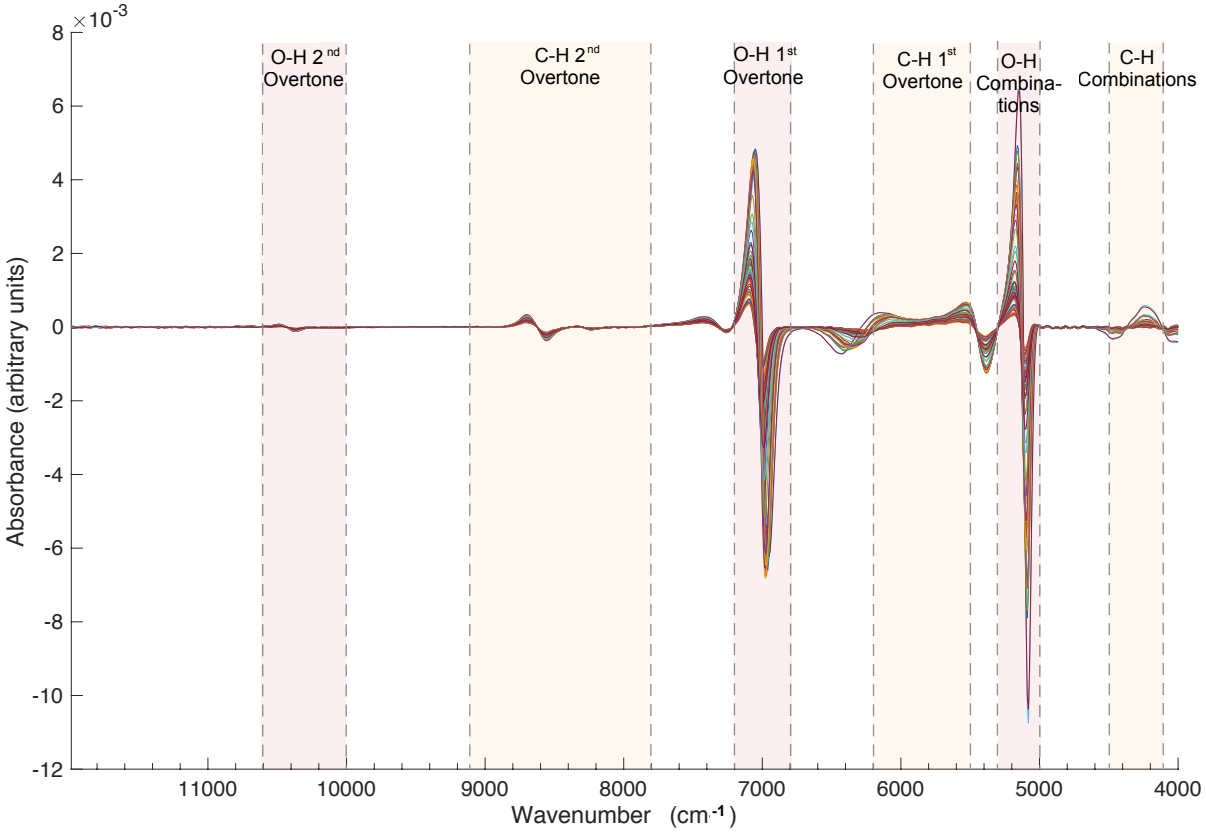
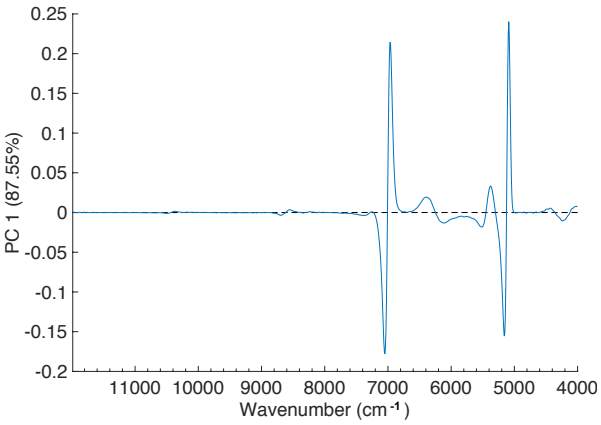


Figure 4.10: Second derivative NIR spectra from at-line monitoring of the accumulation bioreactor during Operation 2 and absorption bands of relevant functional groups in the NIR region. Red indicates the O-H stretching absorption bands and orange refers to the C-H stretching absorption regions. Adapted from [166]

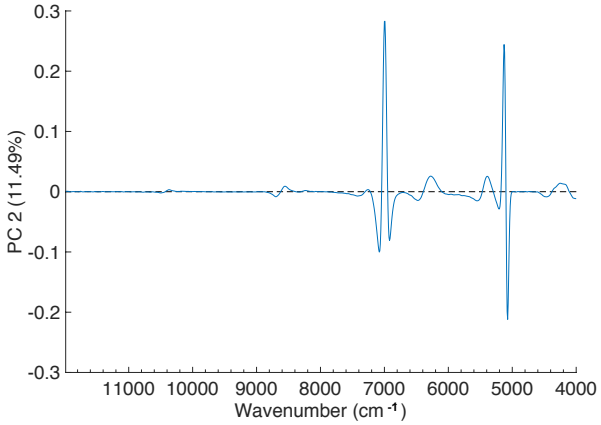
The combinations and first overtone of water (O-H stretching) are very easily identified, with a strong positive peak followed by a major negative peak around 5,100 and 7,000 cm^{-1}, respectively. This was expected, since these regions were already prominent in the raw spectra. However, it is interesting to see that a pair of very subtle positive and negative peaks seems to be present around 10,500 cm^{-1}, which might correspond to the second O-H overtone. Regarding the functional groups of interest, these mostly come down to aliphatic hydrocarbons, since the effects of any carboxylic or amine groups are completely masked by water absorption bands. Thus, only three more spectral regions were identified: the C-H stretching combinations (4,500-4,100 cm^{-1}), its first overtone (6,150-5,500 cm^{-1}) and second overtone (9,100-7,800 cm^{-1}). All three of these intervals comprehend fine peaks with variable intensity, indicating that they might contain useful chemical information regarding the intracellular PHA polymer.

4.3.3 Principal Component Analysis

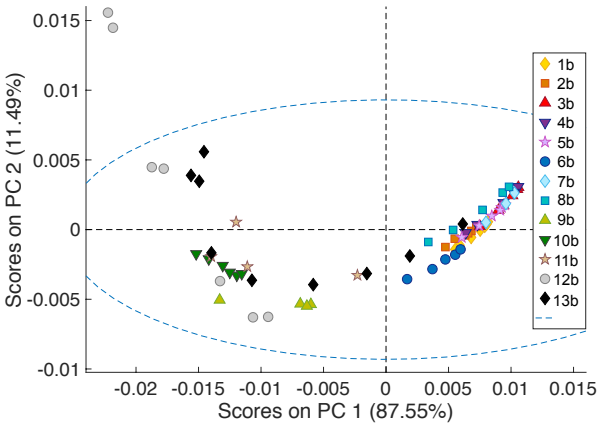
Just like for Raman spectra, a PCA was performed for better data visualization and before moving on to a more quantitative sort of analysis. Primarily, only the second derivative was used as preprocessing and the whole spectral region was investigated, since the goal was to get a broad notion of data trends, sample relationships and variable influence. Again, all samples were further mean-centered.



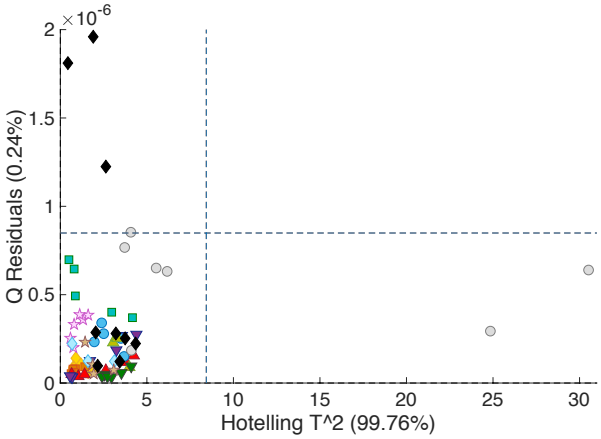
(a) Loadings on PC1.



(b) Loadings on PC2.



(c) Scores plot: PC2 against PC1.



(d) Q residuals against Hotelling T^2 statistics.

Figure 4.11: PCA model on 2nd derivative NIR spectra obtained from the accumulation reactor during Operation 2. The dashed blue line represents a 95% confidence limit.

In this case, two PCs were able to account for 99.04% of the variance in the data. Figure 4.11

exhibits the most relevant plots of the PCA model, including the loadings of both PC1 (a) and PC2 (b), the scores plot (c) and the Q residuals vs Hotelling T^2 plot (d). The scores plot reveals that the first accumulations (1b to 8b), which correspond to the period of higher cell densities prior to changing the C:N ratio, group on the first and fourth quadrants, with positive scores on PC1. The remaining accumulations (9b to 13b), comprehending lower cell densities, score negatively on PC1 and seem to spread somewhat randomly across the second and third quadrants, mostly due to the scores on PC2. In fact, from this second set of accumulations, only the samples of accumulation 10b group together, in clear contrast to those of accumulations 12b and 13b, which are particularly scattered.

Figure 4.11 (d) shows that accumulations 12b and 13b perform badly in terms of the Hotelling T^2 and Q residuals, respectively, further emphasizing their lack of fit in the PCA model. It should be highlighted that these two assays were the only ones in the data set which made use of real fermentate solely, thus implying higher feed volumes and, consequently, more accentuated changes in the mixed liquor matrix. It is possibly the case that not enough comparable samples were introduced to the model and, thus, it lacks robustness regarding accumulations with non-supplemented feed.

Regarding the model loadings, depicted in Figure 4.11 (a) and (b), it is clear that they strongly relate to the second derivative spectra and are both dominated by the water absorption peaks, with some subtle influence of the C-H regions. This validates that the model is indirectly focusing on biomass variation. In fact, the water peaks might serve as an indirect meaure for biomass, since the higher the cell density of the sample, the lower the water absorption will be.

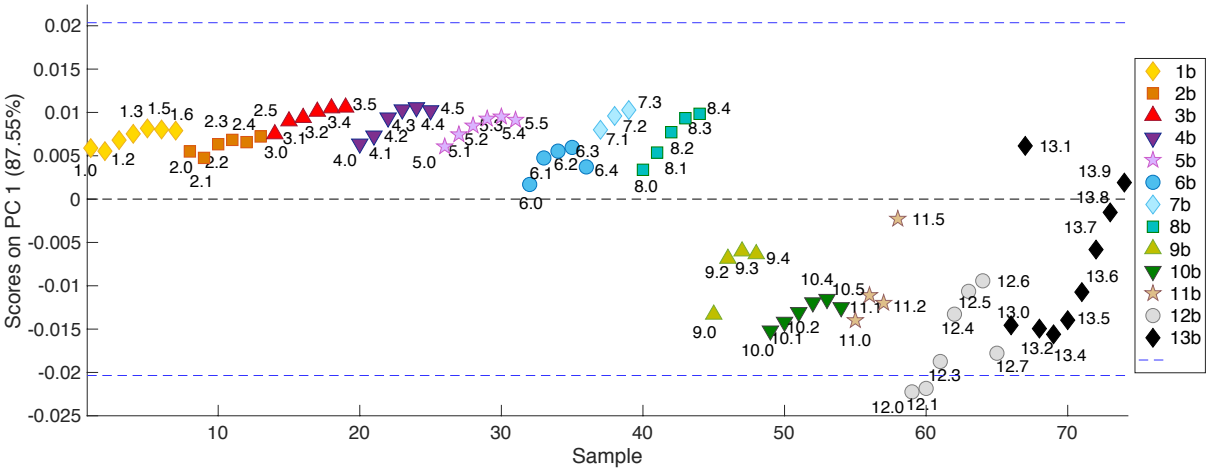


Figure 4.12: PC1 scores from the PCA model on NIR spectra obtained from at-line monitoring of the accumulation bioreactor during Operation 2: PC1 against sample number. The dashed blue line represents a 95% confidence limit.

A closer look at the scores across PC1 can be taken on Figure 4.12. Apart from the previously mentioned segregation between high and low cell densities, it is now possible to detect some trend among each of the accumulations. In fact, it seems to be the general case that samples increase their PC1 score along the accumulations, stressing out, once more, the hypothesis that the data is mostly being compared in terms of its biomass concentration. It is worth noticing, however, that accumulation 12b and 13b cover a considerable range across PC1, with the latter reaching positive PC1 scores, comparable to those of the first accumulations. It is believed that this is not because of a significant cell density increase, which is not supported by the analytical data, but because of a considerable change in the coloration of the

mixed liquor. Indeed, the feed exhibited a very dark color as a result of pH ajustment, and considering the feed volume and the number of pulses, it inevitably led to the gradual blackening of the mixed liquor. It is admitted that the darker the sample the higher the absorption, making it comparable with high cell density samples. Sample 13b-1 was considered an outlier, since no justification was found for its overly high PC1 score and its Q residual had previously been found above the confidence limit.

Before moving on to the subsequent quantitative model development, the PCA results were further used to decide on which accumulation assays to leave out as an external validation set. It was determined that two assays would be excluded, one corresponding to the first period with high cell densities and the other to the latter period after biomass concentration dropped. Based on the scores plot, accumulations 5b and 10b were chosen for this purpose.

4.3.4 Partial Least Squares Regression

PLS is an extension of the PCA method. While PCA extracts factors (PCs) describing the greatest amount of variance in the predictor (X) variables, PLS attempts to find factors (LVs) that both capture variance and achieve correlation with the predicted (Y) variables. In other words, during PLS regression, the spectral information is compared with analytical reference values, so that changes in both data point structures are recognized and correlated.

Given that one of the most useful decision tools for process monitoring is the knowledge of real-time PHA content in the biomass, two different strategies were considered for the development of a quantitative PLS model. The first option was to build a model capable of directly predicting the PHA content (%), whereas the second option was based on building two separate models, one for TSS concentration (g/L) and another for PHA concentration (g/L), and combining these two to obtain PHA content values. These three sets of reference data can be found in Appendix A and each was applied independently in the model development stage. The quality of this reference data is of central importance, since poor chemometric models often result from insufficient accuracy of the reference technique.

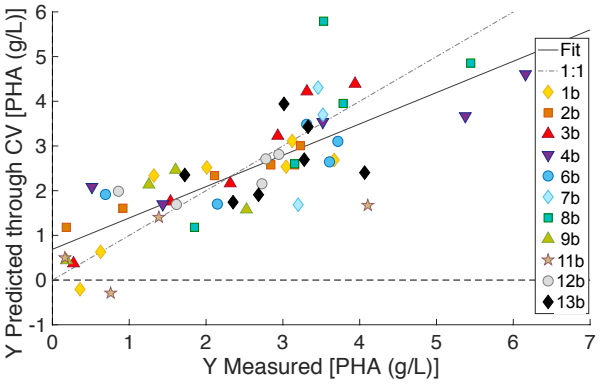
Preliminary Model Development

Following the PCA model, a first PLS model for PHA concentration (g/L) was built, under the same preprocessing conditions: second derivative and mean-centering were applied and the whole spectral region was used. The only difference relied on the removal of accumulations 5b and 10b, along with a few extra samples, from the calibration set. Data points 12b-0 and 12b-1 were considered outliers, based on their unusal scores on the previous PCA model, and so was the previously mentioned sample 13b-1. On the other hand, sample 4b-4 was excluded based on its extremely high PHA concentration, which placed it very far from the other data points. Lastly, sample 13b-0 was also removed, after noticing that this led to a better fitted model. Hence, a total of 57 data points composed the calibration spectral data block. The following model mostly aims to exemplify the several steps and strategies applied during the development of a PLS regression model. The most relevant plots are illustrated in Figure 4.13.

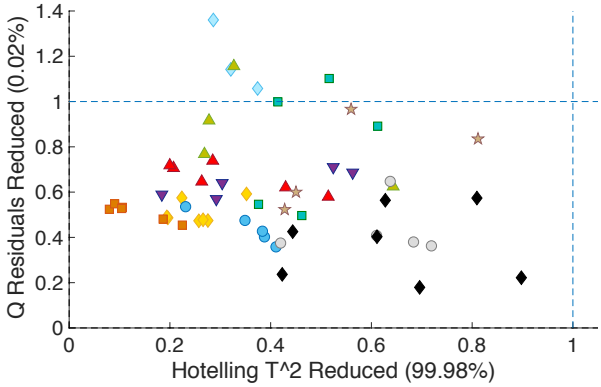
After selecting the spectral data set and the preprocessing method, the selection of the appropriate number of factors follows. Too few factors will lead to a poor explanation of the changes and correlations in the data, while too many may lead to an ”overfitting” scenario, in which the model ends up accounting for the smallest changes in the data set (e.g. spectral noise). The optimal number of LVs was determined based on the minimum of the RMSECV. Thus, it should be highlighted that before moving on to a test set validation, the calibration models were validated with the PLS toolbox default cross-validation method

(venetian blinds with 8 splits and 1 sample per split). Figure 4.13 (c) shows that the minimal RMSECV was obtained at 12 LVs, which was then chosen as the optimal number of LVs. This was also the number of LVs that yielded the highest cross-validation coefficient of determination (R^2 CV), as depicted in Figure 4.13 (d).

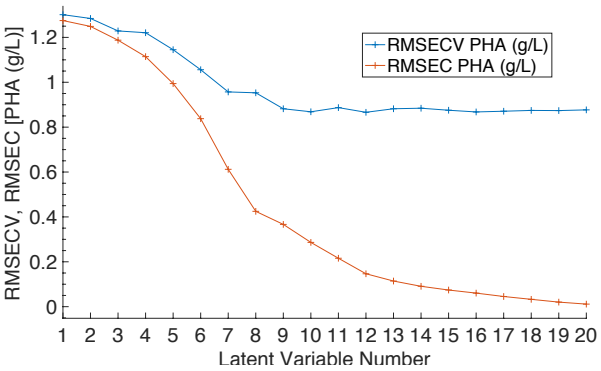
With 12 LVs, the model was able to capture 99.98% and 98.86% of the X-block and Y-block variance, respectively. It also presented a RMSEC of 0.147 g/L and a calibration R^2 of 0.989. Despite these promising calibration parameters, the low cross-validation R^2 value of 0.616 and the substantial RMSECV of 0.8660 g/L obtained imply that the present model has a poor prediction capacity and is not able to extract the relevant information from the spectral data. This is evidentiated by Figure 4.13 (a), which shows the discrepancy between the slope of the fitted curve and the ideal slope of 1, as well as the scattering of the data points around this curve. However, there do not seem to be any strong outliers, judging by Figure 4.13 (b), and thus no further samples were excluded from the data set.



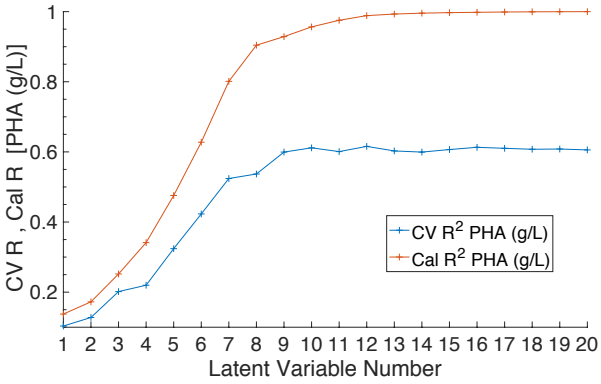
(a) Predicted against measured PHA concentration (g/L).



(b) Reduced Q residuals against reduced Hotelling T^2.



(c) RMSEC and RMSECV against LV number



(d) Calibration and cross-validation R^2 vs LV number.

Figure 4.13: PLS model calibration results for PHA concentration (g/L) based on the 2nd derivative NIR spectra of 57 samples collected from the accumulation bioreactor during Operation 2. Preprocessing was based on second derivatization and mean-centering. An optimal LV of 12 was chosen, capturing 99.98% of X-block variance and 98.86 of Y-block variance. The dashed blue line represents a 95% confidence limit.

The results suggest that further optimization should be seeked, since the prediction performance of the present model is far from satisfactory. The same was noted for the preliminary models for TSS concentration and PHA content, and thus the following section will focus on exploring different optimization strategies for each of the models.

Model Optimization

The first step towards model optimization was to introduce and explore new preprocessing options, looking to enhance the spectral signal. Considering the preprocessing techniques mentioned in Subsection 1.5.3 and their corresponding effects, five different methods were studied: SNV, first derivative, second derivative, first derivative followed by SNV and second derivative followed by SNV.

These preprocessing strategies were further combined with spectral truncation. In fact, the selection of spectral ranges is often of central importance, since the absorption bands of interfering components, for instance, will be accounted for by the PLS algorithm, deteriorating the quality of the model and its subsequent results. Four different spectral truncations were specifically selected for the TSS concentration model and for the PHA content and PHA concentration models. Hence, by combining both the preprocessing options and the spectral ranges, a total of 20 models were developed and evaluated for each predicted variable.

Regarding the models for PHA prediction, both content and concentration, the four spectral ranges selected include the whole spectrum, the C-H stretching first overtone (6150-5500 cm^{-1}), the first and second overtones simultaneously (6150-5500 cm^{-1}+9100-7800 cm^{-1}) and, lastly, both overtones together with the combinations region (6150-5500 cm^{-1}+9100-7800 cm^{-1}+4500-4100 cm^{-1}). Tables 4.4 and 4.5 summarize the most relevant details of all the models developed for PHA concentration and content, respectively. Each model was mainly evaluated in terms of the following parameters: RMSEC, RMSECV, calibration R^2 and cross-validation R^2. This analysis was visually aided by applying a color gradient to each of the columns, in which dark green indicates the best performances (i.e. highest R^2 and lowest error) and red refers to the worst results (i.e. lowest R^2 and highest error) within the 20 models.

Model	Preprocessing	Spectral Regions (cm^{-1})	%VarX	%VarY	#LVs	RMSEC	RMSECV	R^2 cal	R^2 CV
1	SNV	Global	100	88,7	13	0,4629	0,7291	0,8870	0,7449
2	SNV	6150-5500	99,99	79,32	6	0,6364	0,7775	0,7932	0,6922
3	SNV	6150-5500+9100-7800	100	95,74	13	0,2835	0,7859	0,9574	0,6964
4	SNV	6150-5500+9100-7800+4500-4100	100	95,41	14	0,2942	0,8972	0,9641	0,6271
5	1st derivative	Global	100	98,73	15	0,1541	0,6678	0,9873	0,7701
6	1st derivative	6150-5500	100	89,79	11	0,4416	0,7237	0,8979	0,7337
7	1st derivative	6150-5500+9100-7800	100	97,28	12	0,2264	0,7077	0,9728	0,7452
8	1st derivative	6150-5500+9100-7800+4500-4100	100	97,63	13	0,2112	0,6626	0,9763	0,7746
9	2nd derivative	Global	99,98	98,86	12	0,1467	0,8662	0,9886	0,6157
10	2nd derivative	6150-5500	99,96	88,84	11	0,4587	0,8465	0,8884	0,6493
11	2nd derivative	6150-5500+9100-7800	99,86	97,63	10	0,2114	0,6410	0,9763	0,7851
12	2nd derivative	6150-5500+9100-7800+4500-4100	99,88	99,97	19	0,0227	0,4753	0,9997	0,8832
13	1st der+SNV	Global	99,97	95,63	13	0,2870	0,8173	0,9563	0,6681
14	1st der+SNV	6150-5500	99,96	88,66	10	0,4624	0,8079	0,8866	0,6707
15	1st der+SNV	6150-5500+9100-7800	99,98	99,48	18	0,0994	0,7179	0,9948	0,7579
16	1st der+SNV	6150-5500+9100-7800+4500-4100	99,93	95,77	11	0,2835	0,7459	0,9577	0,7131
17	2nd der+SNV	Global	99,72	93,03	10	0,3626	0,8926	0,9303	0,5878
18	2nd der+SNV	6150-5500	98,22	57,07	4	0,8996	1,0804	0,5707	0,3898
19	2nd der+SNV	6150-5500+9100-7800	99,1	97,26	10	0,2271	0,6221	0,9726	0,7969
20	2nd der+SNV	6150-5500+9100-7800+4500-4100	98,94	98,96	13	0,1406	0,5558	0,9896	0,8394

Table 4.4: Models for PHA concentration (g/L) with corresponding parameters and results: preprocessing strategy selected, spectral regions included, optimal number of LVs, % variance captured from the X-block and Y-block, calibration and cross-validation RMSE (g/L) obtained and calibration and cross-validation R^2 achieved. The color gradient indicates the worst (red) to best (green) models in terms of calibration and cross-validation RMSE and R^2

By inspection of Table 4.4, it is clear that differentiation is essential to extract relevant spectral information, since all models relying on SNV exclusively (model 1 to 4) had poor calibration and prediction performances. Overall, it can also be concluded that when only applying first or second derivatization, the model improves with the addition of further C-H stretching absoption bands, i.e, the best results are obtained when all three regions are selected (model 8 and 12). Furthermore, applying SNV after differentiation (model 13 to 20) did not seem to improve the models when comparing with differentiation alone (model 5 to 12). In fact, implementing a SNV normalization after differentiation seems to deteriorate the signal, with model 18 showing particularly bad results. On the other hand, the best performance is achieved with model 12, although it should be noted that the number of LVs is significantly high and this might not lead to equally good results when applying an external validation set. It is also worth highlighting that model 9 corresponds to the PLS model exemplified in the previous section.

Ultimately, models 8, 11, 12, 19 and 20 were chosen as the five most promising models for predicting PHA concentration, to be used in the subsequent test set validation. The selection was mostly based on the RMSECV and cross-validation R^2, while still considering the calibration parameters.

Table 4.5: Models for PHA content (%) with corresponding parameters and results: preprocessing strategy selected, spectral regions included, optimal number of LVs, % variance captured from the X-block and Y-block, calibration and cross-validation RMSE (%) obtained and calibration and cross-validation R^2 achieved. The color gradient indicates the worst (red) to best (green) models in terms of the calibration and cross-validation RMSE and R^2.

Model	Preprocessing	Spectral Regions (cm^{-1})	%VarX	%VarY	#LVs	RMSEC	RMSECV	R^2 cal	R^2 CV
21	SNV	Global	100	85,52	13	0,0842	0,1341	0,8552	0,6827
22	SNV	6150-5500	99,98	70,04	10	0,1203	0,1744	0,7003	0,4385
23	SNV	6150-5500+9100-7800	100	90,25	13	0,0686	0,1430	0,9025	0,6227
24	SNV	6150-5500+9100-7800+4500-4100	100	96,14	15	0,0432	0,1455	0,9614	0,6304
25	1st derivative	Global	100	99,24	16	0,0186	0,0924	0,9924	0,8133
26	1st derivative	6150-5500	100	84,32	10	0,0870	0,1354	0,8432	0,6357
27	1st derivative	6150-5500+9100-7800	100	95,54	12	0,0464	0,1240	0,9554	0,7035
28	1st derivative	6150-5500+9100-7800+4500-4100	99,99	96,74	12	0,0397	0,0914	0,9674	0,8296
29	2nd derivative	Global	99,98	98,73	12	0,0247	0,1284	0,9873	0,6657
30	2nd derivative	6150-5500	99,96	86,07	11	0,0820	0,1472	0,8607	0,5850
31	2nd derivative	6150-5500+9100-7800	99,83	94,17	8	0,0530	0,1134	0,9419	0,7419
32	2nd derivative	6150-5500+9100-7800+4500-4100	99,86	99,78	17	0,0103	0,0915	0,9978	0,8270
33	1st der+SNV	Global	99,96	91,1	12	0,0656	0,1452	0,9110	0,6086
34	1st der+SNV	6150-5500	99,95	79,59	9	0,0993	0,1623	0,7959	0,5115
35	1st der+SNV	6150-5500+9100-7800	99,95	80,06	8	0,0981	0,1483	0,8006	0,5835
36	1st der+SNV	6150-5500+9100-7800+4500-4100	99,93	93,5	11	0,0560	0,1159	0,9350	0,7398
37	2nd der+SNV	Global	99,67	84,54	9	0,0864	0,1552	0,8454	0,5236
38	2nd der+SNV	6150-5500	98,17	46,72	4	0,1632	0,2078	0,4672	0,2137
39	2nd der+SNV	6150-5500+9100-7800	98,36	79,87	5	0,0867	0,1301	0,8444	0,6588
40	2nd der+SNV	6150-5500+9100-7800+4500-4100	99,03	98,25	14	0,0290	0,1194	0,9825	0,7164

Regarding the models for PHA content, the trends are very similar to those found for the previous PHA concentration models: normalization alone does not perform well, better results are obtained for differentiation alone than for differentiation followed by normalization, with second derivative plus SNV (model 38) performing particularly bad. On the other hand, the first derivative models proved to be specially adequate for predicting PHA content, with model 25 achieving some of the best results while using the entire spectra. Finally, models 25, 28, 31, 32 and 36 were selected to be further validated.

Moving on to the optimization of the TSS concentration model, the selected spectral regions were significantly different from the previous ones. Apart from the whole spectral range, the region between 9000 and 7200 cm^{-1} was one of the studied options, to evaluate how the baseline shift in that region was indicative of TSS concentration changes. Furthermore, the water absorption peaks (7200-6800 cm^{-1}+5300-5000 cm^{-1}) were also chosen for the same purpose, since they might serve as an indirect measure of biomass concentration. Lastly, the C-H stretching regions (6150-5500 cm^{-1}+9100-7800 cm^{-1}+4500-4100 cm^{-1}) were also one of the selected truncation solutions, since biomass comprises a significant fraction of organic matter. Just like for PHA concentration and content, these four spectral ranges were combined with the five preprocessing strategies and yielded twenty different models, which are detailed in Table 4.6

Table 4.6: Models for TSS concentration (g/L) with corresponding parameters and results: preprocessing strategy selected, spectral regions included, optimal number of LVs, % variance captured from the X-block and Y-block, calibration and cross-validation RMSE (g/L) obtained and calibration and cross-validation R^2 achieved. The color gradient indicates the worst (red) to best (green) models in terms of the calibration and cross-validation RMSE and R^2.

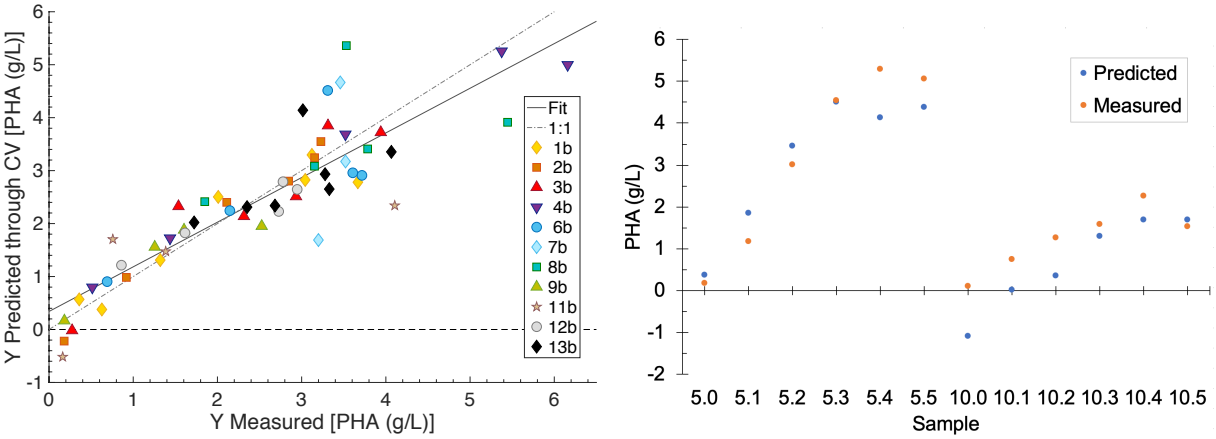
Model	Preprocessing	Spectral Regions (cm^{-1})	%VarX	%VarY	#LVs	RMSEC	RMSECV	R^2 cal	R^2 CV
41	SNV	Global	100	93,74	13	0,4894	0,8908	0,9384	0,8008
42	SNV	9000-7200	100	96,24	10	0,3795	0,8175	0,9624	0,8281
43	SNV	6150-5500+9100-7800+4500-4100	100	93,94	11	0,4816	0,8196	0,9394	0,8278
44	SNV	7200-6800+5300-5000	100	96,6	13	0,3607	0,7199	0,9660	0,8723
45	1st derivative	Global	100	98,48	13	0,2423	0,6958	0,9848	0,8788
46	1st derivative	9000-7200	100	96,84	10	0,3479	0,8311	0,9684	0,8211
47	1st derivative	6150-5500+9100-7800+4500-4100	99,99	93,52	8	0,4980	0,6683	0,9352	0,8839
48	1st derivative	7200-6800+5300-5000	100	98,52	19	0,2377	0,6688	0,9852	0,8919
49	2nd derivative	Global	99,97	95,49	7	0,4152	0,6955	0,9549	0,8755
50	2nd derivative	9000-7200	98,77	90,63	4	0,5988	0,7583	0,9063	0,8506
51	2nd derivative	6150-5500+9100-7800+4500-4100	99,61	95,4	7	0,4194	0,7320	0,9540	0,8616
52	2nd derivative	7200-6800+5300-5000	100	98,48	19	0,2408	0,6631	0,9848	0,8979
53	1st der+SNV	Global	99,94	96,71	9	0,3548	0,6266	0,9668	0,8974
54	1st der+SNV	9000-7200	99,43	96,2	8	0,3780	0,8402	0,9620	0,8205
55	1st der+SNV	6150-5500+9100-7800+4500-4100	99,84	92,33	7	0,5501	0,7744	0,9223	0,8473
56	1st der+SNV	7200-6800+5300-5000	99,99	95,06	11	0,4349	0,7050	0,9506	0,8730
57	2nd der+SNV	Global	99,68	96,25	8	0,3789	0,5947	0,9625	0,9080
58	2nd der+SNV	9000-7200	77,4	92,2	3	0,5414	0,7553	0,9220	0,8492
59	2nd der+SNV	6150-5500+9100-7800+4500-4100	96,93	91,48	8	0,2924	0,6668	0,9783	0,8883
60	2nd der+SNV	7200-6800+5300-5000	99,99	96,73	11	0,3539	0,6012	0,9673	0,9065

The best performances are achieved when selecting the water absorption intervals, followed by the global spectra. Overall, the baseline shift does not seem to be the best predictor for TSS variation. Regarding the preprocessing options, no major conclusions can be drawn, except that SNV alone does not perform very well. It should be noted, however, that the results obtained for TSS concentration models are all considerably positive and the cells colored in red or orange do not necessarily imply poor performance parameters, just worse in comparison with the rest. In fact, all models exhibit RMSECV values below 0.9 g/L and cross-validation R^2 above 0.8. In the end, models 48, 52, 53, 57 and 60 were selected for subsequent validation.

Test Set Validation

In the final phase of model development, accumulations 5b and 10b, which had been previously excluded from the calibration block, were used as an independent set for validation. Prediction was carried out for all 5 models selected for each variable and the best model was chosen based on the minimal RMSEP. Figures 4.14 to 4.16 illustrate the predictive capacity of the final models for PHA concentration (g/L), PHA content (%) and TSS concentration (g/L), respectively. The figures depict both the cross-validation performance, by plotting the predicted values against the measured ones, and the external validation outcome, by plotting the predicted values, together with the measured ones, against the samples in the test set. This latter representation allows for the visualization of the trend of predicted values along the accumulation.

The final PHA concentration model (model 8) refers to a preprocessing comprising the first derivative alone, and uses all three C-H stretching absorption regions (6150-5500 cm^{-1}+9100-7800 cm^{-1}+4500-4100 cm^{-1})). Very good calibration was achieved, with R^2 of 0.976 and a RMSEC of 0.211 g/L. Cross-validation, on the other hand, yielded R^2 of 0.775 and a RMSECV of 0.663 g/L, which are certainly worse than those of calibration but are still satisfactory. Lastly, the test set validation led to R^2 of 0.888 and a RMSEP of 0.693 g/L. The RMSE of cross-validation and prediction are comparable and the prediction R^2 is actually higher than that of cross-validation. This is probably because the desired trend can be found across the predicted values, but these values are slightly underestimated, especially for accumulation 10b. Overall, the model seems to perform better for accumulation 5b and is particularly inappropriate for very low PHA concentrations, predicting a negative concentration for the first sample in accumulation 10b. In fact, the model is likely to be poor at predicting low PHA concentration, namely those obtained after the first few accumulation pulses, given that the RMSEP is considerably high. However, it should provide a more or less significant prediction of the final accumulation pulses. It should be noted that the PHA concentration range in the calibration data set was between 0.16 g/L and 6.16 g/L, which means that sample 10b-0 (0.10 g/L) was extrapolated.

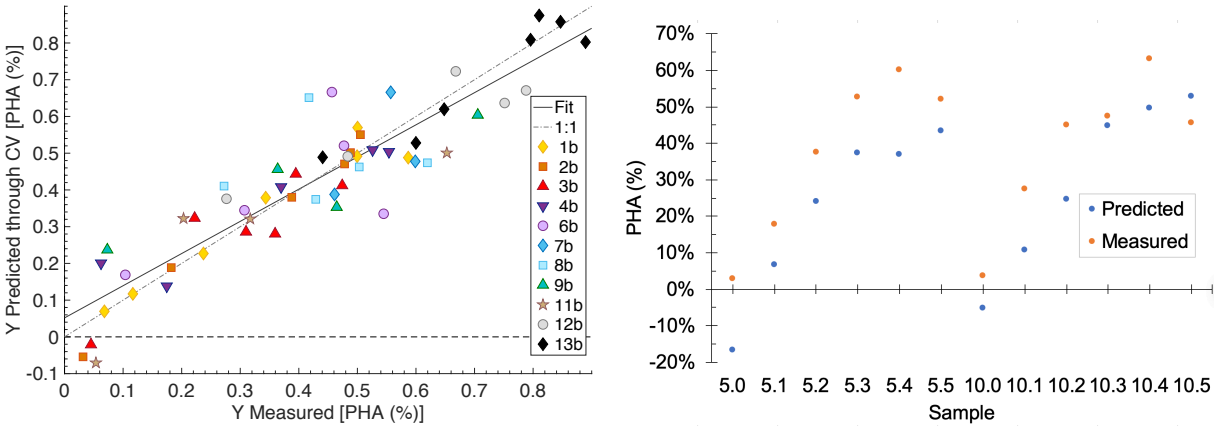


(a) Cross-validation prediction against measured value. **(b)** Predicted and measured values vs sample number.

Figure 4.14: Prediction performance of the final model for PHA concentration (g/L) - model 8. Samples 5.0 to 5.5 and 10.0 to 10.5 in (b) refer to accumulations 5b and 10b, respectively.

Regarding the PHA content prediction, the final model (model 28) also refers to a preprocessing strategy based on the first derivative alone, and to a spectral truncation including all three C-H stretching absorption regions (6150-5500 cm^{-1}+9100-7800 cm^{-1}+4500-4100 cm^{-1}). Again, very good calibration was achieved, with R^2 of 0.967 and a RMSEC of 3.97%. Cross-validation, in turn, led to R^2 of 0.830 and

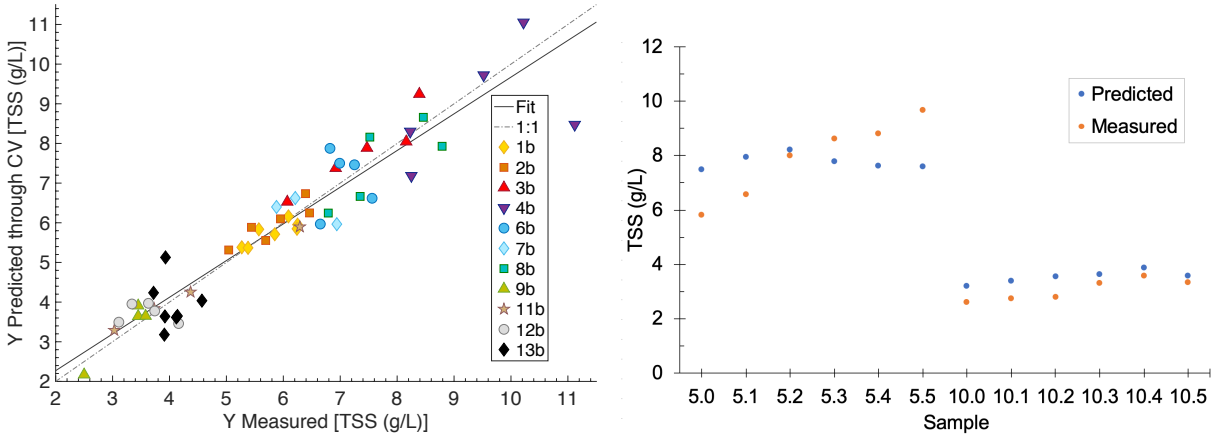
a RMSECV of 9.14%, which are rather promising. Lastly, the test set validation yielded R^2 of 0.864 and a RMSEP of 14.55%. Again, the prediction R^2 is higher than that of cross-validation and the underlying reason is probably the same. In this case, the underestimation is more accentuated and negative PHA concentrations are predicted for the first samples of both accumulations. However, it should be noted that sample 5b-0 (2.95%) is slighly outside the PHA content range of the calibration set, which covers from 3.16% to 88.95%.



(a) Cross-validation prediction against measured value. **(b)** Predicted and measured values vs sample number.

Figure 4.15: Prediction performance of the final model for PHA content (%) - model 28.

Likewise, the final TSS concentration model (model 53) refers to a preprocessing strategy based on the first derivative followed by SNV normalization, and made use of the whole spectral range. The calibration yielded a R^2 of 0.967 and a RMSEC of 0.355 g/L. Cross-validation, in turn, led to a R^2 of 0.897 and a RMSECV of 0.627 g/L, which are very satisfactory results. Lastly, the test set validation resulted in a R^2 of 0.863 and a RMSEP of 1.029 g/L, still fairly decent. Unlike with the previous two models, there does not seem to be any underestimation (nor overestimation) of the concentration values and, in this case, accumulation 10b seems slightly better adjusted. However, since there are no major variations in the TSS concentration values of the test set samples, it is not clear how well the model would react to significant changes and trends in the data. It should be noted that the TSS concentration values in the calibration set ranged between 2.50 g/L and 11.12 g/L and that the reference values of the test set are all within this interval.



(a) Cross-validation prediction against measured value. **(b)** Predicted and measured values vs sample number.

Figure 4.16: Prediction performance of the final model for TSS concentration (g/L) - model 53.

Lastly, since it was stipulated that one of the most relevant parameters to be monitored was the PHA content in the biomass, a final scenario was evaluated, regarding the combination of the final models for PHA and TSS concentration. Hence, the PHA concentration values predicted by model 8 were divided by the TSS concentration values predicted by model 53 and the resulting PHA content predictions were also plotted against the test set samples, along with the true measured values (see Figure 4.17). The final model was surprisingly accurate at predicting the PHA content in accumulation 5b, although a considerable underestimation was again noticed in accumulation 10b. It is probably the case that the model performs better at predicting the samples from high cell density accumulations because a higher number of samples belonging to this period was included in the calibration set. A RMSEP of 18.46% was determined for this combined model. However, this value is almost exclusively the result of predicting accumulation 10b. In fact, a RMSEP of 4.52% and R^2 of 0.960 are achieved if considering accumulation 5b alone.

Nevertheless, the results are extremely promising and clearly confirm the applicability of NIR spectroscopy as a monitoring tool for PHA concentration, PHA content and TSS concentration, particularly in a pilot-scale 3-stage process with MMCs and fruit pulp waste as a feedstock.

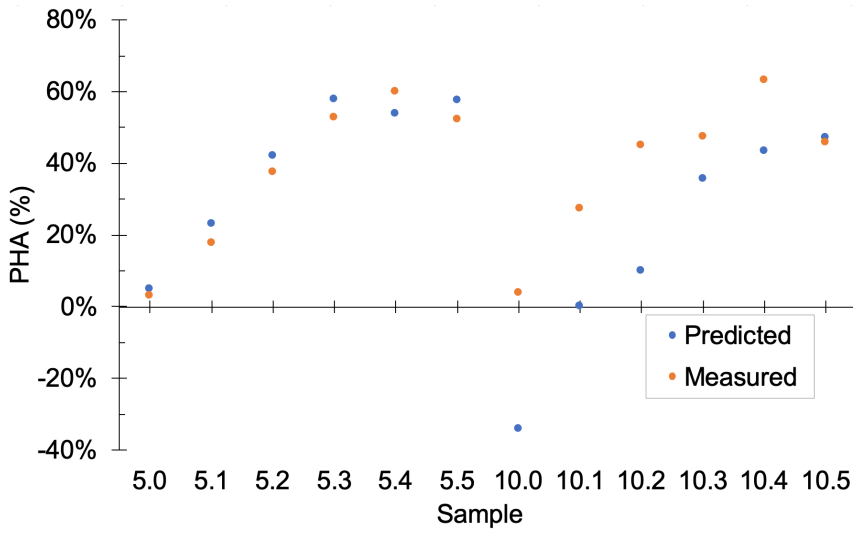


Figure 4.17: PHA content (%) predicted through the combination of model 8 and model 53 and measured values against sample number in test set. Samples 5.0 to 5.5 and 10.0 to 10.5 refer to accumulations 5b and 10b, respectively.

5 Conclusions and Future Work

The pilot-scale production of PHA with MMC using fruit pulp waste as feedstock was successfully demonstrated. A microbial culture with high PHA accumulation capacity was selected in two different operational periods. A FISH analysis carried out during the second operation indicated the presence of genera Azoarcus, Thauera and Plasticicumulans, which are well-known for their PHA-accumulating capacity. The subsequent accumulation assays led to very positive results, with PHA contents above 70% and cell densities up to 17 g/L. The HV content was most frequently between 20% and 30%, although some extreme cases were observed, with HV reaching up to 50% in the final polymer, mostly due to some inconsistencies in the feeding strategy.

Regarding the application of Raman spectroscopy in process monitoring, the present thesis was unable to answer most of the initially raised questions. Unfortunately, difficulties regarding the reference analytical techniques prevented the development of a quantitative model and thus the suitability of the technique for measuring PHA could not be inferred. Nevertheless, a qualitative analysis of 95 Raman spectra, corresponding to 16 accumulation assays, revealed some interesting shifts in the shape of the signal over the weeks. This was believed to be caused by changes in the bacterial community, which was not yet stable by the end of the operation period. Since no FISH analyses were carried out during this time, it was also not possible to corroborate this hypothesis. Furthermore, a PCA model was developed, which clearly separated the samples into three groups according to the shape of the spectra. A rising trend in PC1 scores was observed within the accumulations belonging to the last group, pointing out some potential regarding a quantitative analysis. Lastly, a few subtle peaks were found to be transversal to all spectra, regardless of their main shape, which might also indicate the presence of relevant chemical information. Hence, future work should fall upon gathering accurate analytical information on PHA content, in order to develop a quantitative model and conclude about the applicability of Raman spectroscopy in this context. It might be necessary to perform a rigorous spectral truncation, given the apparent susceptibility of the equipment to changes in the culture. It is also suggested that FISH analyses are performed on a regular basis in order to conclude about this issue. However, it should be kept in mind that the modular Raman equipment available will unlikely lead to a predictive model as successful as those reported in the literature, which make use of more powerful techniques like micro-Raman spectroscopy [128, 129].

On the contrary, NIR spectroscopy showed great promise for the monitoring of the PHA production process. A total of 74 samples, with their corresponding reference values, was gathered. Even though the NIR spectra were dominated by strong water absorption bands, it was still possible to identify the C-H stretching overtones and combinations. A PCA model using the second derivative spectra revealed some sample grouping based on biomass concentration which essentially separated the data in two sets of accumulations: before and after decreasing the nitrogen levels in the SBR feed. As for the quantitative analysis, three different approaches were explored: a model for predicting PHA content (%) directly,

and the combination of two models for predicting PHA and TSS concentration (g/L), separately. Model optimization was carried out by investigating different preprocessing strategies and several spectral truncations. Apart from cross-validation, external validation was performed on the most promising models for each predicted variable. The test set comprised two accumulation assays which had been previously excluded from the calibration set, one from each of the above mentioned PCA groups. The best models for prediction of PHA concentration made use of all three C-H stretching absorption regions (6150-5500 cm^{-1}+9100-7800 cm^{-1}+4500-4100 cm^{-1}) with first derivative alone as preprocessing and yielded a RMSEP of 0.69 g/L and R^2 of 0.89. The chosen model for PHA content used the same spectral regions and preprocessing and led to a RMSEP of 14.6% and R^2 of 0.86. The final TSS concentration model, on the other hand, made use of the whole spectral range and its preprocessing involved the first derivative followed by SNV normalization, which resulted in a RMSEP of 1.0 g/L and R^2 of 0.86. Overall, the models performed better at predicting the accumulation belonging to the high cell density group (before decreasing nitrogen levels in the SBR feed), since more calibration points were available regarding this period. Finally, the model combining the PHA and TSS concentration models performed rather poorly for the second accumulation in the test set but revealed extraordinary results when considering the high cell density accumulation only, with a RMSEP of 4.52% and a R^2 of 0.96.

It is clear that NIR spectroscopy seems a very promising tool for predicting PHA concentration, PHA content and TSS concentration in the present production process. It is believed that the RMSEPs obtained can be significantly improved by gathering more samples in the calibration set, because it seems like the model is simply not robust enought at this point. It might also be the case that the linear relationship between the NIR signal and the analyte concentration depends on the TSS concentration range and then two different models would have to be developed for higher and lower cell densities. On the other hand, the analytical methods still present some accuracy issues and might be hindering the application of the NIR technique to its full potential. Finally, it might also be interesting to further investigate the prediction capacity for each monomer individually, as perhaps the NIR technology might be able to predict the HB:HV ratio in the final copolymer.

Ultimately, once a more robust and reliable calibration model has been developed, the next step should be to insert the probe inside the accumulation reactor to test the applicability of the model in an online situation. This might require further calibration but, once established, will enable real-time measurements and, consequently, a more comprehensive study of the production dynamics. It should be noted, however, that regular reference analytical data will always be needed in order to guarantee the long-term validity of the method.

As far the author is aware, this is the first time that a NIR-based quantitative model has been reported for the monitoring of a pilot-scale production of P(3HB-co-HV) polymer using MMC and a complex feedstock as substrate.

Bibliography

[1] Martin Koller and Gerhart Braunegg. “Advanced approaches to produce polyhydroxyalkanoate (PHA) biopolyesters in a sustainable and economic fashion”. In: *The EuroBiotech Journal* 2.2 (2018), pp. 89–103. DOI: 10.2478/ebtj-2018-0013.

[2] PlasticsEurope. *Plastics - the facts 2018*. Tech. rep. 2018. DOI: 10.1016/j.marpolbul.2013.01.015. arXiv: arXiv:1011.1669v3.

[3] Laurent Lebreton and Anthony Andrady. “Future scenarios of global plastic waste generation and disposal”. In: *Palgrave Communications* 5.1 (2019), pp. 1–11. ISSN: 2055-1045. DOI: 10.1057/s41599-018-0212-7. URL: http://dx.doi.org/10.1057/s41599-018-0212-7.

[4] Ellen MacArthur Foundation. *The New Plastics Economy*. Tech. rep. Ellen MacArthur Foundation, 2016, pp. 1–120. DOI: 10.1080/13563460500344468.

[5] Ritchie Hannah Roser, Max. *Plastic Pollution*. 2018. URL: https://ourworldindata.org/plastic-pollution (visited on 04/18/2019).

[6] Jenna R Jambeck, Roland Geyer, Chris Wilcox, Theodore R Siegler, Miriam Perryman, Anthony Andrady, Ramani Narayan, and Kara Lavender Law. “Plastic waste inputs from land into the ocean”. In: *American Association for the Advancement of Science* 347.6223 (2015), pp. 768–771. DOI: https://doi.org/10.1126/science.1260352.

[7] European Commission. *A European Strategy for Plastics*. Tech. rep. 2018, p. 24. DOI: 10.1021/acs.est.7b02368. URL: http://ec.europa.eu/environment/circular-economy/pdf/plastics-strategy-brochure.pdf%7B%5C%%7D0Ahttp://ec.europa.eu/environment/circular-economy/index%7B%5C_%7Den.htm.

[8] *‘A line in the sand’ - Ellen MacArthur Foundation launches New Plastics Economy Global Commitment to eliminate plastic waste at source*. 2018. URL: https://www.ellenmacarthurfoundation.org/news/a-line-in-the-sand-ellen-macarthur-foundation-launch-global-commitment-to-eliminate-plastic-pollution-at-the-source.

[9] *New Plastics Economy Global Commitment*. Tech. rep. June. Ellen MacArthur Foundation, 2019.

[10] M. Reis, M. Albuquerque, M. Villano, and M. Majone. “Mixed Culture Processes for Polyhydroxyalkanoate Production from Agro-Industrial Surplus/Wastes as Feedstocks”. In: *Comprehensive Biotechnology, Second Edition*. Vol. 6. Elsevier Inc., Sept. 2011, pp. 669–683. ISBN: 9780080885049. DOI: 10.1016/B978-0-08-088504-9.00464-5.

[11] *Ypack: Reduce, Reuse, Recover*. URL: https://www.ypack.eu/partners/ (visited on 04/22/2019).

[12] *Bio-based Industries Joint Undertaking: a €3.7 billion partnership between the EU and the Bio-based Industries Consortium*. URL: https://www.bbi-europe.eu/projects/fresh (visited on 04/22/2019).

[13] *Polybioskin: high performance functional bio-based polymers for skin-contact products in the biomedical, cosmetic and sanitary industries.* URL: http://polybioskin.eu (visited on 04/22/2019).

[14] *Circpack: From Waste to Resource*. URL: http://circpack.eu (visited on 04/22/2019).

[15] *Glopack: Granting society with LOw environmental impact innovative PACKaging*. URL: https://glopack2020.eu/about/ (visited on 04/22/2019).

[16] Scott Lambert and Martin Wagner. "Environmental performance of bio-based and biodegradable plastics: The road ahead". In: *Chemical Society Reviews* 46.22 (2017), pp. 6855–6871. ISSN: 14604744. DOI: 10.1039/c7cs00149e. URL: http://dx.doi.org/10.1039/C7CS00149E.

[17] Michel Vert, Yoshiharu Doi, Karl-heinz Hellwich, Michael Hess, Philip Hodge, Przemyslaw Kubisa, Marguerite Rinaudo, and François Schué. "Terminology for Biorelated Polymers and Applications". In: *Chemistry International – Newsmagazine for IUPAC* 33.2 (2014), pp. 377–410. ISSN: 0193-6484. DOI: 10.1515/ci.2011.33.2.25c.

[18] European Bioplastics. *Bioplastics. Disponível em: http://en.europeanbioplastics.org/bioplastics/ Acesso em: 11 jul.* Tech. rep. Berlin, 2016. URL: https://docs.european-bioplastics.org/2016/publications/fs/EUBP%7B%5C_%7Dfs%7B%5C_%7Dwhat%7B%5C_%7Dare%7B%5C_%7Dbioplastics.pdf.

[19] T A Cooper and Argo Group International. *Developments in bioplastic materials for packaging food , beverages and other fast- moving consumer goods*. Woodhead Publishing Limited, pp. 108–152. ISBN: 9780857098979. DOI: 10.1533/9780857098979.108. URL: http://dx.doi.org/10.1533/9780857098979.108.

[20] Ana R V Ferreira, Vítor D Alves, and Isabel M Coelhoso. "Polysaccharide-Based Membranes in Food Packaging Applications". In: 2015 (2016), pp. 1–17. DOI: 10.3390/membranes6020022.

[21] Kumar Sudesh and Tadahisa Iwata. "Sustainability of biobased and biodegradable plastics". In: *Clean - Soil, Air, Water* 36.5-6 (2008), pp. 433–442. ISSN: 18630650. DOI: 10.1002/clen.200700183.

[22] Nafisa Jabeen, Ishrat Majid, and Gulzar Ahmad Nayik. "Bioplastics and food packaging: A review". In: *Cogent Food & Agriculture* 1.1 (2015), pp. 1–6. DOI: 10.1080/23311932.2015.1117749.

[23] Xiang Qi, Yiwei Ren, and Xingzu Wang. "New advances in the biodegradation of Poly (lactic) acid". In: *International Biodeterioration & Biodegradation* 117 (2017), pp. 215–223. ISSN: 0964-8305. DOI: 10.1016/j.ibiod.2017.01.010. URL: http://dx.doi.org/10.1016/j.ibiod.2017.01.010.

[24] R Prathipa, C Sivakumar, and B Shanmugasundaram. "Biodegradable polymers for sustainable packaging applications". In: *International Journal of Mechanical Engineering and Technology* 9.6 (2018), pp. 293–303. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85049729331%7B%5C&%7DpartnerID=40%7B%5C&%7Dmd5=c049e1a53e29071069de67db440a0c7a.

[25] E Castro-aguirre, F Iñiguez-franco, H Samsudin, X Fang, and R Auras. "Poly (lactic acid) — Mass production , processing , industrial applications , and end of life". In: *Advanced Drug Delivery Reviews* 107 (2016), pp. 333–366. ISSN: 0169-409X. DOI: 10.1016/j.addr.2016.03.010. URL: http://dx.doi.org/10.1016/j.addr.2016.03.010.

[26] *Biomer*. URL: http://www.biomer.de/IndexE.html (visited on 08/29/2019).

[27] *TianAn Biopolymer - Nature's Ecofriendly Solution*. URL: http://www.tianan-enmat.com/%7B%5C#%7D (visited on 08/29/2019).

[28] *Danimer Scientific - a Biotechnology Company*. URL: https://danimerscientific.com (visited on 08/29/2019).

[29] European Bioplastics. “Facts and Figures”. In: (2018). ISSN: 18347819. DOI: 10.1111/j.1834-7819.1975.tb05057.x. URL: https://docs.european-bioplastics.org/publications/EUBP%7B%5C_%7DFacts%7B%5C_%7Dand%7B%5C_%7Dfigures.pdf.

[30] European Bioplastics. *Global production capacities of bioplastics 2018-2023*. Tech. rep. 2018.

[31] Tajalli Keshavarz and Ipsita Roy. “Polyhydroxyalkanoates : bioplastics with a green agenda”. In: *Current Opinion in Microbiology* 13.3 (2010), pp. 321–326. ISSN: 1369-5274. DOI: 10.1016/j.mib.2010.02.006. URL: http://dx.doi.org/10.1016/j.mib.2010.02.006.

[32] Jingnan Lu, Ryan C Tappel, and Christopher T Nomura. “Mini-Review : Biosynthesis of Poly (hydroxyalkanoates) Mini-Review : Biosynthesis of”. In: *Polymer Review* 49.3 (2009), pp. 226–248. DOI: 10.1080/15583720903048243.

[33] Akhilesh Kumar Singh and Nirupama Mallick. “SCL-LCL-PHA copolymer production by a local isolate , Pseudomonas aeruginosa MTCC 7925”. In: *Biotechnology Journal* 4 (2009), pp. 703–711. DOI: 10.1002/biot.200800307.

[34] K Sudesh, H Abe, and Y Doi. “Synthesis , structure and properties of polyhydroxyalkanoates : biological polyesters”. In: 25 (2000), pp. 1503–1555.

[35] Sevgi Bayarı and Feride Severcan. “FTIR study of biodegradable biopolymers : P (3HB), P (3HB- co -4HB) and P (3HB- co -3HV)”. In: *Journal of Molecular Structure* 747 (2005), pp. 529–534. DOI: 10.1016/j.molstruc.2004.12.029.

[36] Yuanpeng Wang, Ronghui Chen, Jiyuan Cai, Zhenggui Liu, Yanmei Zheng, Haitao Wang, and Qingbiao Li. “Biosynthesis and Thermal Properties of PHBV Produced from Levulinic Acid by Ralstonia eutropha”. In: *PLoS One* 8.4 (2013), pp. 4–11. DOI: 10.1371/journal.pone.0060318.

[37] Constantina Kourmentza, Jersson Pl, Nikolaos Venetsaneas, Anna Burniol-figols, Cristiano Varrone, Hariklia N Gavala, and Maria A M Reis. “Recent Advances and Challenges towards Sustainable Polyhydroxyalkanoate (PHA) Production”. In: (), pp. 1–43. DOI: 10.3390/bioengineering4020055.

[38] Martin Koller, Lukas Marsalek, Miguel Miranda de Sousa Dias, and Gerhart Braunegg. “Producing microbial polyhydroxyalkanoate (PHA) biopolyesters in a sustainable manner”. In: *New Biotechnology* 37 (2017), pp. 24–38.

[39] S Yan, Bala Subramanian S, R D Tyagi, and R Y Surampalli. “Polymer production by bacterial strains isolated from activated sludge treating municipal wastewater”. In: *Water Science Technology* 57.4 (2008), pp. 533–539.

[40] Doan Van-Thuoc, Tran Huu-Phong, Nguyen Thi-Binh, Nguyen Thi-Tho, Duong Minh-Lam, and Jorge Quillaguamán. “Polyester production by halophilic and halotolerant bacterial strains obtained from mangrove soil samples located in Northern Vietnam”. In: *MicrobiologyOpen* 1.4 (2012), pp. 395–406. DOI: https://dx.doi.org/10.1002%2Fmbo3.44.

[41] Rodrigo Yoji Uwamori Takahashi, Nathalia Aparecida Santos Castilho, Marcus Adonai Castro da Silva, Maria Cecilia Miotto, and André Oliveira De Souza Lima. “Prospecting for Marine Bacteria for Polyhydroxyalkanoate Production on Low-cost Substrates”. In: *Bioengineering*. Vol. 4. 3. 2017, p. 60. DOI: 10.3390/bioengineering4030060.

[42] Wang Fengqin and Sang Yup Lee. “Poly(3-Hydroxybutyrate) Production with High Productivity and High Polymer Content by a Fed-Batch Culture of Alcaligenes latus under Nitrogen Limitation.” In: *Applied and environmental microbiology* (1997).

[43] Chad Nielsen, Asif Rahman, Asad Ur Rehman, Marie K. Walsh, and Charles D. Miller. “Food waste conversion to microbial polyhydroxyalkanoates”. In: *Microbial Biotechnology* 10.6 (2017), pp. 1338–1352. ISSN: 17517915. DOI: 10.1111/1751-7915.12776.

[44] M Teresa Cesário, Rodrigo S Raposo, M Catarina MD De Almeida, Frederik Van Keulen, Bruno S Ferreira, and M Manuela R Fonseca. “Enhanced bioproduction of poly-3-hydroxybutyrate from

wheat straw lignocellulosic hydrolysates". In: *New Biotechnology* 31.1 (2014), pp. 104–113. DOI: 10.1016/j.nbt.2013.10.004.

[45] José Geraldo da Cruz Pradella, Jaciane Lutz Ienczak, Cecilia Romero Delgado, and Marilda Keico Taciro. "Carbon source pulsed feeding to attain high yield and high productivity in poly (3-hydroxybutyrate) (PHB) production from soybean oil using Cupriavidus necator". In: *Biotechnology Letters* 34 (2012), pp. 1003–1007. DOI: 10.1007/s10529-012-0863-1.

[46] Fabrício Coutinho de Paula, Sérgio Kakazu, Carolina Bilia Chimello De Paula, José Gregório Cabrera Gomez, and Jonas Contiero. "Polyhydroxyalkanoate production from crude glycerol by newly isolated Pandoraea sp ." In: *Journal of King Saud University – Science* 29 (2017), pp. 166–173. DOI: 10.1016/j.jksus.2016.07.002.

[47] Woo Suk Ahn, Si Jae Park, and Sang Yup Lee. "Production of poly (3-hydroxybutyrate) from whey by cell recycle fed-batch culture of recombinant Escherichia coli". In: *Biotechnology Letters* 23 (2001), pp. 235–240.

[48] Majd Khalid Esthaya, Nor'Aini Abdul Rhaman, and Mohd Ali Hassan. "Bioconversion of restaurant waste into Polyhydroxybutyrate (PHB) by recombinant E . coli through anaerobic digestion Majd Khalid Eshtaya , Nor ' Aini Abdul Rahman and Mohd Ali Hassan". In: *International Journal of Environment and Waste Management* 11.1 (2013), pp. 27–37.

[49] C.-C. Chien and L.-Y. Ho. "Polyhydroxyalkanoates production from carbohydrates by a genetic recombinant Aeromonas sp ." In: *Letters in Applied Microbiology* 47 (2008), pp. 587–593. DOI: 10.1111/j.1472-765X.2008.02471.x.

[50] Silvana Povolo, Paolo Toffano, Marina Basaglia, and Sergio Casella. "Bioresource Technology Polyhydroxyalkanoates production by engineered Cupriavidus necator from waste material containing lactose". In: *Bioresource Technology* 101.20 (2010), pp. 7902–7907. ISSN: 0960-8524. DOI: 10.1016/j.biortech.2010.05.029. URL: http://dx.doi.org/10.1016/j.biortech.2010.05.029.

[51] Guo-qiang Chen and Xiao-ran Jiang. "Engineering bacteria for enhanced polyhydroxyalkanoates (PHA) biosynthesis". In: *Synthetic and Systems Biotechnology* 2.3 (2017), pp. 192–197. ISSN: 2405-805X. DOI: 10.1016/j.synbio.2017.09.001. URL: https://doi.org/10.1016/j.synbio.2017.09.001.

[52] Qian Liu, Ge Luo, Xin Rong, and Guo-qiang Chen. "Biosynthesis of poly (3-hydroxydecanoate) and 3-hydroxydodecanoatedominating polyhydroxyalkanoates by b -oxidation pathway inhibited Pseudomonas putida". In: *Metabolic Engineering* 13.1 (2011), pp. 11–17. ISSN: 1096-7176. DOI: 10.1016/j.ymben.2010.10.004. URL: http://dx.doi.org/10.1016/j.ymben.2010.10.004.

[53] Lakshmi Tripathi, Lin-ping Wu, Meng Dechuan, Jinchun Chen, Qiong Wu, and Guo-qiang Chen. "Bioresource Technology Pseudomonas putida KT2442 as a platform for the biosynthesis of polyhydroxyalkanoates with adjustable monomer contents and compositions". In: *Bioresource Technology* 142 (2013), pp. 225–231. ISSN: 0960-8524. DOI: 10.1016/j.biortech.2013.05.027. URL: http://dx.doi.org/10.1016/j.biortech.2013.05.027.

[54] Xiao-ran Jiang and Guo-qiang Chen. "Morphology engineering of bacteria for bio-production". In: *Biotechnology Advances* (2015). ISSN: 0734-9750. DOI: 10.1016/j.biotechadv.2015.12.007. URL: http://dx.doi.org/10.1016/j.biotechadv.2015.12.007.

[55] Dan Tan, Qiong Wu, Jin-chun Chen, and Guo-qiang Chen. "Engineering Halomonas TD01 for the low-cost production of polyhydroxyalkanoates". In: *Metabolic Engineering* 26 (2014), pp. 34–47. ISSN: 1096-7176. DOI: 10.1016/j.ymben.2014.09.001. URL: http://dx.doi.org/10.1016/j.ymben.2014.09.001.

[56] Sulfiqar Ali Raza, Sharjeel Abid, and Ibrahim M Banat. “Polyhydroxyalkanoates : Characteristics , production , recent developments and applications”. In: *International Biodeterioration & Biodegradation* 126.October 2017 (2018), pp. 45–56. DOI: 10.1016/j.ibiod.2017.10.001.

[57] Lei Pei and Markus Schmidt. “Fast-Growing Engineered Microbes : New Concerns for Gain-of-Function”. In: 9.June (2018), pp. 1–7. DOI: 10.3389/fgene.2018.00207.

[58] Bronwyn Laycock, Peter Halley, Steven Pratt, Alan Werker, and Paul Lant. “The chemomechanical properties of microbial polyhydroxyalkanoates”. In: *Progress in Polymer Science* 38.3-4 (2013), pp. 536–583. ISSN: 0079-6700. DOI: 10.1016/j.progpolymsci.2013.06.008. URL: http://dx.doi.org/10.1016/j.progpolymsci.2013.06.008.

[59] Luisa S Serafim, Paulo C Lemos, Maria G E Albuquerque, and Maria A M Reis. “Strategies for PHA production by mixed cultures and renewable waste materials”. In: *Applied Microbiology and Biotechnology* 81 (2008), pp. 615–628. DOI: 10.1007/s00253-008-1757-y.

[60] Xiaofei Wang, Adrian Oehmen, Elisabete B Freitas, Gilda Carvalho, and Maria A M Reis. “The link of feast-phase dissolved oxygen (DO) with substrate competition and microbial selection in PHA production”. In: *Water Research* 112 (2017), pp. 269–278. ISSN: 0043-1354. DOI: 10.1016/j.watres.2017.01.064. URL: http://dx.doi.org/10.1016/j.watres.2017.01.064.

[61] Sebastian L Riedel, Jingnan Lu, Ulf Stahl, and Christopher J Brigham. “Lipid and fatty acid metabolism in Ralstonia eutropha : relevance for the biotechnological production of value-added products”. In: *Applied Microbiology and Biotechnology* 98 (2014), pp. 1469–1483. DOI: 10.1007/s00253-013-5430-8.

[62] Filipa Pardelha, Maria G E Albuquerque, Maria A M Reis, João M L Dias, and Rui Oliveira. “Flux balance analysis of mixed microbial cultures : Application to the production of polyhydroxyalkanoates from complex mixtures of volatile fatty acids”. In: *Journal of Biotechnology* 162 (2012), pp. 336–345. ISSN: 0168-1656. DOI: 10.1016/j.jbiotec.2012.08.017. URL: http://dx.doi.org/10.1016/j.jbiotec.2012.08.017.

[63] Filipa Pardelha, Maria G E Albuquerque, Maria A M Reis, and M L Dias. “Dynamic metabolic modelling of volatile fatty acids conversion to polyhydroxyalkanoates by a mixed microbial culture”. In: *New Biotechnology* 31.4 (2014), pp. 335–344. DOI: 10.1016/j.nbt.2013.06.008.

[64] Francesco Valentino, Fernando Morgan-Sagastume, Sabrina Campanari, Marianna Villano, Alan Werker, and Mauro Majone. “Carbon recovery from wastewater through bioconversion into biodegradable polymers”. In: *New BIOTECHNOLOGY* 25.37 (2017), pp. 9–23. ISSN: 1871-6784. DOI: 10.1016/j.nbt.2016.05.007. URL: http://dx.doi.org/10.1016/j.nbt.2016.05.007.

[65] H Takabatake, H Satoh, T Mino, and T Matsuo. “Recovery of biodegradable plastics from activated sludge process”. In: *Water Science Technology* 42.3-4 (2000), pp. 351–356.

[66] M G E Albuquerque, V Martino, E Pollet, L Avérous, and M A M Reis. “Mixed culture polyhydroxyalkanoate (PHA) production from volatile fatty acid (VFA) -rich streams : Effect of substrate composition and feeding regime on PHA productivity , composition and properties”. In: *Journal of Biotechnology* 151.1 (2011), pp. 66–76. ISSN: 0168-1656. DOI: 10.1016/j.jbiotec.2010.10.070. URL: http://dx.doi.org/10.1016/j.jbiotec.2010.10.070.

[67] Markets and Markets. *Polyhydroxyalkanoate (PHA) Market by Type (Monomers, Co-Polymers, Terpolymers), Manufacturing Technology (Bacterial Fermentation, Biosynthesis, Enzymatic Catalysis), Application (Packaging, Bio Medical, Food Services, Agriculture) - Global Forecast to 202*. Tech. rep. 2017.

[68] *Danimer Scientific*. URL: https://danimerscientific.com/about-us/history/ (visited on 08/30/2019).

[69] *Danimer Scientific*. URL: https://danimerscientific.com/about-us/partnerships/uga/ (visited on 08/30/2019).

[70] *CISION PR Newswire*. URL: https://www.prnewswire.com/news-releases/danimer-scientific-creates-first-fully-biodegradable-plastic-straw-300709557.html (visited on 08/30/2019).

[71] *Danimer Scientific*. URL: https://danimerscientific.com/about-us/partnerships/pepsi/ (visited on 08/30/2019).

[72] Muhammadi, Shabina, Muhammad Afzal, and Shafqat Hameed. "Bacterial polyhydroxyalkanoates-eco-friendly next generation plastic : Production , biocompatibility , biodegradation , physical properties and applications". In: *Green Chemistry Letters and Reviews* 8.3-4 (2015), pp. 56–77. DOI: 10.1080/17518253.2015.1109715.

[73] Martin Koller. "Biodegradable and Biocompatible Polyhydroxy-alkanoates (PHA): Auspicious Microbial Macromolecules for Pharmaceutical and Therapeutic Applications". In: *Molecules* 23.2 (2018). DOI: 10.3390/molecules23020362.

[74] Janice Lim, Mingliang You, Jian Li, and Zibiao Li. "Emerging bone tissue engineering via Polyhydroxyalkanoate (PHA)-based scaffolds". In: *Materials Science and Engineering: C* 79 (2017), pp. 917–929. DOI: https://doi.org/10.1016/j.msec.2017.05.132.

[75] Kai Zhao, Ying Deng, Jin Chun Chen, and Guo-qiang Chen. "Polyhydroxyalkanoate (PHA) scaffolds with good mechanical properties and biocompatibility". In: *Biomaterials* 24 (2003), pp. 1041–1045.

[76] Zhang Junyu, Ekaterina I Shishatskaya, Tatiana G Volova, Luiziana Ferreira, and Guo-qiang Chen. "Polyhydroxyalkanoates (PHA) for therapeutic applications". In: *Materials Science & Engineering C* 86 (2018), pp. 144–150. ISSN: 0928-4931. DOI: 10.1016/j.msec.2017.12.035. URL: https://doi.org/10.1016/j.msec.2017.12.035.

[77] Colin W Pouton and Saghir Akhtarb. "Biosynthetic polyhydroxyalkanoates and their potential in drug delivery". In: *Advanced Drug Delivery Reviews* 18 (1996), pp. 133–162.

[78] Verena Peters and Bernd H A Rehm. "In Vivo Enzyme Immobilization by Use of Engineered Polyhydroxyalkanoate Synthase". In: *Applied and Environmental Microbiology* 72.3 (2006), pp. 1777–1783. DOI: 10.1128/AEM.72.3.1777.

[79] Abhilasha Singh Mathuriya and J V Yakhmi. "Polyhydroxyalkanoates : Biodegradable Plastics and Their Applications". In: *Handbook of Eco-Materials*. 2017. ISBN: 9783319482811.

[80] Guozhan Jiang, David J Hill, Marek Kowalczuk, Brian Johnston, Grazyna Adamus, Victor Irorere, and Iza Radecka. "Carbon Sources for Polyhydroxyalkanoates and an Integrated Biorefinery". In: *International Journal of Molecular Science* (2016). DOI: 10.3390/ijms17071157.

[81] Martin Koller, Aid Atlic, Miguel Dias, Angelika Reiterer, and Gerhart Braunegg. "Microbial PHA Production from Waste Raw Materials". In: *Plastics from Bacteria*. Vol. 14. Springer, Berlin, Heidelberg, 2010, pp. 85–119. DOI: 10.1007/978-3-642-03287.

[82] Karolin Dietrich, Marie-josée Dumont, Luis F Del, and Valérie Orsat. "Sustainable PHA production in integrated lignocellulose biore fi neries". In: *New Biotechnology* 49 (2019), pp. 161–168. DOI: 10.1016/j.nbt.2018.11.004.

[83] Anouk F Duque, Catarina S S Oliveira, Inês T D Carmo, Ana R Gouveia, Filipa Pardelha, Ana M Ramos, and Maria A M Reis. "Response of a three-stage process for PHA production by mixed microbial cultures to feedstock shift : impact on polymer composition". In: *New Biotechnology* 31.4 (2014), pp. 276–288. DOI: 10.1016/j.nbt.2013.10.010.

[84] F Silva, L Serafim, H Nadais, L Arroja, and I Capela. *Acidogenic Fermentation Towards Valorisation of Organic Waste Streams into Volatile Fatty Acids*. 2013. DOI: 10.1016/0022-0728(95)04447-7.

[85] H Yesil, A E Tugtas, A Bayrakdar, and B Calli. “Anaerobic fermentation of organic solid wastes : volatile fatty acid production and separation”. In: *Water Science & Technology* 69.10 (2014), pp. 2132–2138. DOI: 10.2166/wst.2014.132.

[86] Tasneem Abbasi and S.A. Abbasi. “Formation and impact of granules in fostering clean energy production and wastewater treatment in upflow anaerobic sludge blanket (UASB) reactors”. In: *Renewable and Sustainable Energy Reviews* 16 (2012), pp. 1696–1708.

[87] H Salehizadeh and M C M Van Loosdrecht. “Production of polyhydroxyalkanoates by mixed culture : recent trends and biotechnological importance”. In: *Biotechnology Advances* 22 (2004), pp. 261–279. DOI: 10.1016/j.biotechadv.2003.09.003.

[88] Paulo C Lemos, S Serafim, and Maria A M Reis. “Synthesis of polyhydroxyalkanoates from different short-chain fatty acids by mixed cultures submitted to aerobic dynamic feeding”. In: *Journal of Biotechnology* 122 (2006), pp. 226–238. DOI: 10.1016/j.jbiotec.2005.09.006.

[89] Simon Bengtsson, Jakob Hallquist, Alan Werker, and Thomas Welander. “Acidogenic fermentation of industrial wastewaters : Effects of chemostat retention time and pH on volatile fatty acids production”. In: *Biochemical Engineering Journal* 40 (2008), pp. 492–499. DOI: 10.1016/j.bej.2008.02.004.

[90] M G E Albuquerque, M Eiroa, C Torres, B R Nunes, and M A M Reis. “Strategies for the development of a side stream process for polyhydroxyalkanoate (PHA) production from sugar cane molasses”. In: *Journal of Biotechnology* 130.4 (2007), pp. 411–421. DOI: 10.1016/j.jbiotec.2007.05.011.

[91] Erik R. Coats, Frank J. Loge, William A. Smith, David N. Thompson, and Michael P. Wolcott. “Functional Stability of a Mixed Microbial Consortium Producing PHA From Waste Carbon Sources”. In: *Applied Biochemistry and Biotechnology* 137.1-12 (2007), pp. 909–925.

[92] Catarina S.S. Oliveira, Carlos E. Silva, Gilda Carvalho, and Maria A. Reis. “Strategies for efficiently selecting PHA producing mixed microbial cultures using complex feedstocks: Feast and famine regime and uncoupled carbon and nitrogen availabilities”. In: *New Biotechnology* 37 (2017), pp. 69–79. ISSN: 18764347. DOI: 10.1016/j.nbt.2016.10.008. URL: http://dx.doi.org/10.1016/j.nbt.2016.10.008.

[93] M G E Albuquerque, C A V Torres, and M A M Reis. “Polyhydroxyalkanoate (PHA) production by a mixed microbial culture using sugar molasses : Effect of the influent substrate concentration on culture selection”. In: *Water Research* 44 (2010), pp. 3419–3433. ISSN: 0043-1354. DOI: 10.1016/j.watres.2010.03.021. URL: http://dx.doi.org/10.1016/j.watres.2010.03.021.

[94] Hector García Martín, Natalia Ivanova, Victor Kunin, Falk Warnecke, Kerrie W. Barry, Alice C Mchardy, Christine Yeates, Shaomei He, Asaf A Salamov, Ernest Szeto, Eileen Dalin, Nik H Putnam, Harris J Shapiro, Jasmyn L Pangilinan, Isidore Rigoutsos, Nikos C Kyrpides, Linda Louise Blackall, Katherine D Mcmahon, and Philip Hugenholtz. “Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities”. In: *Nature* 24.10 (2006), pp. 1263–1269. DOI: 10.1038/nbt1247.

[95] Michael Rodgers and Guangxue Wu. “Bioresource Technology Production of polyhydroxybutyrate by activated sludge performing enhanced biological phosphorus removal”. In: *Bioresource Technology* 101.3 (2010), pp. 1049–1053. ISSN: 0960-8524. DOI: 10.1016/j.biortech.2009.08.107. URL: http://dx.doi.org/10.1016/j.biortech.2009.08.107.

[96] Francisco Cabrera, Álvaro Torres, José Luís Campos, and David Jeison. “Effect of Operational Conditions on the Behaviour and Associated Costs of Mixed Microbial Cultures for PHA Production”. In: *Polymers* 11.2 (2019), p. 191. DOI: 10.3390/polym11020191.

[97] J J Beun, F Paletta, M C M Van Loosdrecht, and J J Heijnen. “Stoichiometry and Kinetics of Poly-beta -Hydroxybutyrate Metabolism in Aerobic , Slow Growing , Activated Sludge Cultures”. In: *Biotechnology and Bioengineering* 67.4 (2000), pp. 379–389.

[98] Zhiqiang Chen, Long Huang, Qinxue Wen, Huichao Zhang, and Zirui Guo. “Effects of sludge retention time , carbon and initial biomass concentrations on selection process : From activated sludge to polyhydroxyalkanoate accumulating cultures”. In: *Journal of Environmental Sciences* 52 (2017), pp. 76–84. ISSN: 1001-0742. DOI: 10.1016/j.jes.2016.03.014. URL: http://dx.doi.org/10.1016/j.jes.2016.03.014.

[99] Adeline S M Chua, Hiroo Takabatake, Hiroyasu Satoh, and Takashi Mino. “Production of polyhydroxyalkanoates (PHA) by activated sludge treating municipal wastewater : effect of pH , sludge retention time (SRT), and acetate concentration in influent”. In: *Water Research* 37 (2003), pp. 3602–3611. DOI: 10.1016/S0043-1354(03)00252-5.

[100] Adrian Oehmen, Vera Silva, and Maria A M Reis. “The impact of pH control on the volumetric productivity of mixed culture PHA production from fermented molasses”. In: *Engineering in Life Sciences* 14.2 (2014), pp. 143–152. DOI: 10.1002/elsc.201200220.

[101] Luísa S Serafim, Paulo C Lemos, Rui Oliveira, and Maria A M Reis. “Optimization of Polyhydroxybutyrate Production by Mixed Cultures Submitted to Aerobic Dynamic Feeding Conditions”. In: *Biotechnology and Bioengineering* 87.2 (2004), pp. 145–160. DOI: 10.1002/bit.20085.

[102] Bertan Basak, Orhan Ince, Nazik Artan, Nevin Yagci, and Bahar Kasapgil Ince. “Effect of nitrogen limitation on enrichment of activated sludge for PHA production”. In: *Bioprocess and Biosystems Engineering* 34 (2011), pp. 1007–1016. DOI: 10.1007/s00449-011-0551-x.

[103] Fernando Ramos Silva. “Impact of carbon / nitrogen feeding strategy on polyhydroxyalkanoates production using mixed microbial cultures”. PhD thesis. Universidade Nova de Lisboa, 2015.

[104] André Freches and Paulo C Lemos. “Microbial selection strategies for polyhydroxyalkanoates production from crude glycerol : Effect of OLR and cycle length”. In: *New Biotechnology* 39.A (2017), pp. 22–28. ISSN: 1871-6784. DOI: 10.1016/j.nbt.2017.05.011. URL: http://dx.doi.org/10.1016/j.nbt.2017.05.011.

[105] Hong Chen, Huijuan Meng, Zuchao Nie, and Mingmei Zhang. “Polyhydroxyalkanoate production from fermented volatile fatty acids : Effect of pH and feeding regimes”. In: *Bioresource Technology* 128 (2013), pp. 533–538. ISSN: 0960-8524. DOI: 10.1016/j.biortech.2012.10.121. URL: http://dx.doi.org/10.1016/j.biortech.2012.10.121.

[106] Zhiqiang Chen, Long Huang, Qinxue Wen, and Zirui Guo. “Efficient polyhydroxyalkanoate (PHA) accumulation by a new continuous feeding mode in three-stage mixed microbial culture (MMC) PHA production process”. In: *Journal of Biotechnology* 209 (2015), pp. 68–75. ISSN: 0168-1656. DOI: 10.1016/j.jbiotec.2015.06.382. URL: http://dx.doi.org/10.1016/j.jbiotec.2015.06.382.

[107] Warren Blunt, David B Levin, and Nazim Cicek. “Bioreactor Operating Strategies for Improved Polyhydroxyalkanoate (PHA) Productivity”. In: *Polymers* 10.11 (2018), p. 1197. DOI: 10.3390/polym10111197.

[108] Francesco Valentino, Lamija Karabegovic, Mauro Majone, Fernando Morgan-, and Alan Werker. “Polyhydroxyalkanoate (PHA) storage within a mixed-culture biomass with simultaneous growth as a function of accumulation substrate nitrogen and phosphorus levels.” In: *Water Research* 77 (2015), pp. 49–63. ISSN: 0043-1354. DOI: 10.1016/j.watres.2015.03.016. URL: http://dx.doi.org/10.1016/j.watres.2015.03.016.

[109] Long Huang, Zhiqiang Chen, Qinxue Wen, and Duu-jong Lee. “Enhanced polyhydroxyalkanoate production bymixed microbial culturewithextended cultivationstrategy”. In: *Bioresource Technol-*

ogy 241 (2017), pp. 802–811. ISSN: 0960-8524. DOI: 10.1016/j.biortech.2017.05.192. URL: http://dx.doi.org/10.1016/j.biortech.2017.05.192.

[110] Partha Chakravarty, Vasant Mhaisalkar, and Tapan Chakrabarti. “Bioresource Technology Study on poly-hydroxyalkanoate (PHA) production in pilot scale continuous mode wastewater treatment system”. In: *Bioresource Technology* 101.8 (2010), pp. 2896–2899. ISSN: 0960-8524. DOI: 10.1016/j.biortech.2009.11.097. URL: http://dx.doi.org/10.1016/j.biortech.2009.11.097.

[111] Simon Anterrieu, Luca Quadri, Bert Geurkink, Inez Dinkla, Simon Bengtsson, Monica Arcos-hernandez, Tomas Alexandersson, Fernando Morgan-sagastume, Anton Karlsson, Markus Hjort, Lamija Karabegovic, Per Magnusson, Peter Johansson, Magnus Christensson, and Alan Werker. “Integration of biopolymer production with process water treatment at a sugar factory”. In: *New Biotechnology* 31.4 (2014), pp. 308–323. ISSN: 1871-6784. DOI: 10.1016/j.nbt.2013.11.008. URL: http://dx.doi.org/10.1016/j.nbt.2013.11.008.

[112] F Morgan-Sagastume, M Hjort, D Cirne, F Gerardin, S Lacroix, G Gaval, L Karabegovic, T Alexandersson, P Johansson, A Karlsson, S Bengtsson, M V Arcos-Hernández, P Magnusson, and A Werker. “Integrated production of polyhydroxyalkanoates (PHAs) with municipal wastewater and sludge treatment at pilot scale”. In: *Bioresource Technology* 181.C (2015), pp. 78–89. ISSN: 0960-8524. DOI: 10.1016/j.biortech.2015.01.046. URL: http://dx.doi.org/10.1016/j.biortech.2015.01.046.

[113] Simon Bengtsson, Anton Karlsson, Tomas Alexandersson, Luca Quadri, Markus Hjort, Peter Johansson, Fernando Morgan-Sagastume, Simon Anterrieu, Monica Arcos-Hernandez, Lamija Karabegovic, Per Mangusson, and Alan Werker. “A process for polyhydroxyalkanoate (PHA) production from municipal wastewater treatment with biological carbon and nitrogen removal demonstrated at pilot-scale”. In: *New Biotechnology* 35 (2017), pp. 42–53. ISSN: 1871-6784. DOI: 10.1016/j.nbt.2016.11.005. URL: http://dx.doi.org/10.1016/j.nbt.2016.11.005.

[114] Jelmer Tamis, Kätlin Luˇ, Yang Jiang, Mark C M Van Loosdrecht, and Robbert Kleerebezem. “Enrichment of Plasticicumulans acidivorans at pilot-scale for PHA production on industrial wastewater”. In: *Journal of Biotechnology* 192 (2014), pp. 161–169. DOI: 10.1016/j.jbiotec.2014.10.022.

[115] Sang Yup Lee and Jong-Il Choi. “Effect of fermentation performance on the economics of poly(3-hydroxybutyrate) production byAlcaligenes latus”. In: *Polymer Degradation and Stability* 59.1-3 (1998), pp. 387–393.

[116] Cristina Pérez-Rivero, J Pablo López-Gómez, and Ipsita Roy. “A sustainable approach for the downstream processing of bacterial polyhydroxyalkanoates : State-of-the-art and latest developments”. In: *Biochemical Engineering Journal* 150.June (2019), p. 107283. ISSN: 1369-703X. DOI: 10.1016/j.bej.2019.107283. URL: https://doi.org/10.1016/j.bej.2019.107283.

[117] Nik N M N Irdahayu, Kai-Hee Huong, Vigneswari Sevakumaran, and Shantini Kannusamy. “En route to economical eco-friendly solvent system in enhancing sustainable recovery of poly(3 -hydroxybutyrate -co -4 -hydroxybutyrate) copolymer”. In: *Engineering in Life Sciences* 17.9 (2017). DOI: 10.1002/elsc.201600217.

[118] Vasiliki Kachrimanidou, Nikolaos Kopsahelis, Anestis Vlysidis, Seraphim Papanikolaou, Ioannis K. Kookos, Belén Monje Martínez, María Consuelo E. Rondán, and Apostolis A. Koutinas. “Downstream separation of poly(hydroxyalkanoates) using crude enzyme consortia produced via solid state fermentation integrated in a biorefinery concept”. In: *Food and Bioproducts Processing* 100.A (2016), pp. 323–334. ISSN: 0960-3085. DOI: 10.1016/j.fbp.2016.08.002. URL: http://dx.doi.org/10.1016/j.fbp.2016.08.002.

[119] Parisa Hejazi, Ebrahim Vasheghani-Farahani, and Yadollah Yamini. “Supercritical Fluid Disruption of Ralstonia eutropha for Poly (beta-hydroxybutyrate) Recovery”. In: *Biotechnology Progress* 19 (2003), pp. 1519–1523.

[120] Yoong Kit, Leong Pau, and Loke Show. “Economic and environmental analysis of PHAs production process”. In: *Clean Technologies and Environmental Policy* 19.7 (2017), pp. 1941–1953. ISSN: 1618-9558. DOI: 10.1007/s10098-017-1377-2.

[121] N D Lourenço, J A Lopes, C F Almeida, M C Sarraguça, and H M Pinheiro. “Bioreactor monitoring with spectroscopy and chemometrics : a review”. In: *Analytical and Bioanalytical Chemistry* 404 (2012), pp. 1211–1237. DOI: 10.1007/s00216-012-6073-9.

[122] Jing Dai, Erik R Coats, and Armando G Mcdonald. “Multivariate near infrared spectroscopy for predicting polyhydroxybutyrate biosynthesis by mixed microbial consortia cultured on crude glycerol”. In: *Biomass and Bioenergy* 81 (2015), pp. 490–495. ISSN: 0961-9534. DOI: 10.1016/j.biombioe.2015.08.009. URL: http://dx.doi.org/10.1016/j.biombioe.2015.08.009.

[123] Martin Koller and Alejandra Rodríguez-Contreras. “Techniques for tracing PHA-producing organisms and for qualitative and quantitative analysis of intra- and extracellular PHA”. In: *Engineering in Life Sciences* 15 (2015), pp. 558–581. DOI: 10.1002/elsc.201400228.

[124] Palmiro Poltronieri, Valeria Mezzola, and Oscar F D’urso. “PHB Production in Biofermentors Assisted through Biosensor Applications”. In: *Proceedings* 1.4 (2017). DOI: 10.3390/ecsa-3-E014.

[125] Ewen Smith and Geoffrey Dent. *Modern Raman Spectroscopy - A Practical Approach*. Ed. by Ltd John Wiley & Sons. 2004. ISBN: 9780471496687. DOI: 10.1002/0470011831.

[126] Peter J. Larkin. *IR and Raman Spectroscopy - Principles and Spectral Interpretation*. Ed. by Elsevier Inc. 1st. 2011. ISBN: 978-0-12-386984-5 For.

[127] D. D. Le Pevelen. *NIR FT-Raman*. 2016.

[128] Joke De Gelder, Diana Willemse-erix, Maarten J Scholtes, Jorge I Sanchez, Kees Maquelin, Peter Vandenabeele, Patrick De Boever, Gerwin J Puppels, Luc Moens, and Paul De Vos. “Monitoring Poly (3-hydroxybutyrate) Production in Cupriavidus necator DSM 428 (H16) with Raman Spectroscopy”. In: *Analytical Chemistry* 80.6 (2008), pp. 2155–2160. DOI: 10.1021/ac702185d.

[129] Ota Samek, Stanislav Obruca, Martin Siler, Petr Sedlacek, Pavla Benesova, Dan Kucera, Ivana Márova, Jan Jezek, Silva Bernatová, and Pavel Zemánek. “Quantitative Raman Spectroscopy Analysis of Polyhydroxyalkanoates Produced by Cupriavidus necator H16”. In: *Sensors (Basel)* 16.11 (2016). DOI: 10.3390/s16111808.

[130] Celly M S Izumi and Marcia L A Temperini. “Vibrational Spectroscopy FT-Raman investigation of biodegradable polymers : Poly (3-hydroxybutyrate) and poly (3-hydroxybutyrate- co -3-hydroxyvalerate)”. In: *Vibrational Spectroscopy* 54.2 (2010), pp. 127–132. ISSN: 0924-2031. DOI: 10.1016/j.vibspec.2010.07.011. URL: http://dx.doi.org/10.1016/j.vibspec.2010.07.011.

[131] Verena Jost, Matthias Schwarz, and Horst-christian Langowski. “Investigation of the 3-hydroxyvalerate content and degree of crystallinity of P3HB-co-3HV cast films using Raman spectroscopy”. In: *Polymer* 133 (2017), pp. 160–170. ISSN: 0032-3861. DOI: 10.1016/j.polymer.2017.11.026. URL: https://doi.org/10.1016/j.polymer.2017.11.026.

[132] Annapaola Petrosino. “Assessing the potential of spectroscopy for real-time monitoring of bioreactors producing PHA”. PhD thesis. Universita’ di Bologna, 2018.

[133] Brian G Osborne. *Near-infrared Spectroscopy in Food Analysis*. 2006. DOI: 10.1002/9780470027318.a1018.

[134] Matthew Scarff, S Alison Arnold, Linda M Harvey, and Brian McNeil. “Near Infrared Spectroscopy for Bioprocess Monitoring and Control : Current Status and Future Trends”. In: *Critical Reviews in Biotechnology* 26 (2006), pp. 17–39. DOI: 10.1080/07388550500513677.

[135] Celio Pasquini. “Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Applications”. In: *Journal of the Brazilian Chemical Society* 14.2 (2003), pp. 198–219. ISSN: 01035053. DOI: 10.1590/S0103-50532003000200006.

[136] Jan C. J. Bart, Emanuele Gucciardi, and Stefano Cavallaro. “Quality assurance of biolubricants”. In: *Biolubricants: Science and Technology*. Ed. by Stefano Cavallaro. Woodhead Publishing Limited, 2013. Chap. 8. ISBN: 9780857092632.

[137] Ana L Leal, Jorge C Ribeiro, Artur M S Silva, and F G Martins. “Predicting Research and Motor Octane Numbers based on Near Infrared Spectroscopy : Models based on Partial Least Squares Regression and Artificial Neural Networks”. In: *28th European Symposium onComputer Aided Process Engineering*. Vol. 43. Elsevier Masson SAS, 2018, pp. 187–192. ISBN: 9780444642356. DOI: 10.1016/B978-0-444-64235-6.50034-6. URL: https://doi.org/10.1016/B978-0-444-64235-6.50034-6.

[138] Madalena V Cruz, Mafalda Cruz Sarraguça, Filomena Freitas, João Almeida Lopes, and Maria A M Reis. “Online monitoring of P (3HB) produced from used cooking oil with near-infrared spectroscopy”. In: *Journal of Biotechnology* 194 (2015), pp. 1–9. DOI: 10.1016/j.jbiotec.2014.11.022.

[139] Meng-lin Zhang, Guo-ping Sheng, Yang Mu, Wei-hua Li, Han-qing Yu, Hideki Harada, and Yu-You Li. “Rapid and accurate determination of VFAs and ethanol in the effluent of an anaerobic H2 -producing bioreactor using near-infrared spectroscopy”. In: *Water Research* 43.7 (2009), pp. 1823–1830. ISSN: 0043-1354. DOI: 10.1016/j.watres.2009.01.018. URL: http://dx.doi.org/10.1016/j.watres.2009.01.018.

[140] Barry M. Wise, Neal B. Gallagher, Rasmus Bro, Jeremy M. Shaver, Willem Windig, and R. Scott Koch. *PLS_Toolbox 3.5 for use with MATLAB*. Eigenvector Research, Inc, 2005. ISBN: 0-9761184-0-8.

[141] Tom Dearing. *Fundamentals of Chemometrics and Modeling*.

[142] Kurt-Erik Haggblom. *Basics of Multivariate Modelling and Data Analysis*.

[143] Rekha Gautam, Sandeep Vanga, Freek Ariese, and Siva Umapathy. “Review of multidimensional data processing approaches for Raman and infrared spectroscopy”. In: *EPJ Techniques and Instrumentation* 2.8 (2015). DOI: 10.1140/epjti/s40485-015-0018-6.

[144] Bart M Nicolai, Katrien Beullens, Els Bobelyn, Ann Peirs, Wouter Saeys, Karen I Theron, and Jeroen Lammertyn. “Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy : A review”. In: *Postharvest Biology and Technology* 46 (2007), pp. 99–118. DOI: 10.1016/j.postharvbio.2007.06.024.

[145] Kristian Hovde Liland and Nils Kristian Afseth. “Model-based pre-processing in Raman spectroscopy of biological samples”. In: *Journal of Raman Spectroscopy* 47.6 (2016), pp. 643–650. DOI: 10.1002/jrs.4886.

[146] J Huang, S Romero-Torres, and M Moshgbar. “Practical Considerations in Data Pre-treatment for NIR and Raman Spectroscopy”. In: *American Pharmaceutical Review* 13.6 (2010), pp. 116–127.

[147] Yves Roggo, Pascal Chalus, Lene Maurer, Carmen Lema-Martinez, Aurélie Edmond, and Nadine Jent. “A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies”. In: *Journal of Pharmaceutical and Biomedical Analysis* 44 (2007), pp. 683–700. DOI: 10.1016/j.jpba.2007.03.023.

[148] *PCA basics*. 2014.

[149] Kevin A. James, Peter J. Woolf, and Shayn M. Peirce. “High-level Modeling of Biological Networks”. In: *Systems Biomedicine: Concepts and Perspectives*. Academic Press, 2010. Chap. 9, pp. 225–247.

[150] American Public Health Association, American Water Works Association, and Water Environment Federation. *Standard Methods for the Examination of Water and Wastewater*. Ed. by Rodger B. Baird, Andrew D. Eaton, and Eugene W. Rice. 23rd. Washington, 2017, 2540A.

[151] Amann RI. "In situ identification of microorganisms by whole cell hybridization with rRNA-targeted nucleic acid probes." In: *Molecular Microbial Ecology Manual*. Ed. by ADL Akkermans, JD van Elsas, and FJ de Bruijn. Dordrecht: Springer, 1995, pp. 1–15. DOI: https://doi.org/10.1007/978-94-011-0351-0_23.

[152] P.H. Nielsen, H. Daims, and H. Lemmer. *FISH Handbook for Biological Wastewater Treatment: Identification and Quantification of Microorganisms in Activated Sludge and Biofilms by FISH.* London: IWA Publishing Company, 2009.

[153] R.I. Amann, B.J. Binder, R.J. Olson, S.W. Chisholm, R. Devereux, D.A. Stahl, and D.A. "Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations." In: *Applied and Environmental Microbiology* 56 (1990), pp. 1919–1925.

[154] M. Daims, H. Bruhl, A. Amann, R. Schleifer, and K.H. Wagner. "The domain-specific probe EU338 is insufficient for the detection of all Bacteria: development and evaluation of a more comprehensive probe set." In: *Systematic and Applied Microbiology* 22 (1999), pp. 434–444.

[155] A. Neef, A. Zaglauer, H. Meier, R. Amann, H Lemmer, and K.H. Schleifer. "opulation analysis in a denitrifying sand filter: conventional an in situ identification of Paracoccus sp. in methanol-fed biofilms." In: *Applied and Environmental Microbiology* 62 (1996), pp. 4329–4339.

[156] A. Hess, B. Zarda, D. Hahn, A. Haner, D. Stax, P. Hohener, and J. Zeyer. "In situ analysis of denitrifying toluene- and m-xylene-degrading bacteria in a diesel fuel-contaminated laboratory aquifer column." In: *Applied and Environmental Microbiology* 63 (1997), pp. 2136–2141.

[157] A. Loy, C. Schulz, S. Lucker, A. Schopfer-Wendels, K. Stoecker, C. Baranyi, A. Lehner, and M. 2005. Wagner. "16S rRNA gene-based oligonucleotide microarray for environmental monitoring of the betaproteobacterial order Rhodocyclales." In: *Applied and Environmental Microbiology* 71 (2005), pp. 1373–1386.

[158] N. Lee, C.M. Cellamare, C. Bastianutti, R. Rosseló-Mora, P. Kampfer, W. Ludwig, K.H. Schleifer, and L. Stante. "Emended description of the species Lampropedia hyalina." In: *International Journal of Systematic and Evolutionary Microbiology* 54 (2004), pp. 1709–1715.

[159] A Maszenan, RJ Seviour, BK Patel, and J. Wanner. "A fluorescently-labelled r-RNA targeted oligonucleotide probe for the in situ detection of G-bacteria of the genus Amaricoccus in activated sludge." In: *Journal of Applied Microbiology* 88.5 (2000), pp. 826–835. DOI: 10.1046/j.1365-2672.2000.01022.x.

[160] R.A. Rosselló-Mora, M. Wagner, R. Amann, and K.H. Schleifer. "The abundance of Zoogloea ramigera in sewage treatment plants". In: *Applied and Environmental Microbiology* 61 (1995), pp. 702–707.

[161] K. Johnson, Y. Jiang, R. Kleerebezem, G. Muyzer, and M.C. van Loosdrecht. "Enrichment of a mixed bacterial culture with a high polyhydroxyalkanoate storage capacity." In: *Biomacromolecules* 10 (2009), pp. 670–676.

[162] Sijmen De Jong, Barry M Wise, and N Lawrence Ricker. "Canonical partial least squares and continuum power regression". In: *Journal of Chemometrics* 15 (2001), pp. 85–100.

[163] Dieter Naumann. "FT-Infrared and FT-Raman Spectroscopy in Biomedical Research". In: *Applied Spectroscopy Reviews* 36.2-3 (2001), pp. 239–298. DOI: http://dx.doi.org/10.1081/ASR-100106157PLEASE.

[164] R. J. A. do Nascimento, G R de Macedo, E. S. dos Santos, and J A de Oliveira. "Real Time and In Sity Near-Infrared Spectroscopy (NIRS) for Quantitative Monitoring of Biomass, Glucose, Ethanol and Glycerine Concencentration in an Alcoholic Fermentation". In: *Brazilian Journal of*

Chemical Engineering 34.02 (2017), pp. 459–468. DOI: dx.doi.org/10.1590/0104-6632.20170342s20150347ISSN.

[165] J P Reed, D Devlin, S R R Esteves, R Dinsdale, and A J Guwy. “Performance parameter prediction for sewage sludge digesters using reflectance FT-NIR spectroscopy”. In: *Water Research* 45.8 (2011), pp. 2463–2472. ISSN: 0043-1354. DOI: 10.1016/j.watres.2011.01.027. URL: http://dx.doi.org/10.1016/j.watres.2011.01.027.

[166] Jorg-Peter Conzen. *Multivariate Calibration - a Practical Guide for Developing Methods in the Quantitative Analytical Chemistry*. 3rd Englis. Bruker Optik GmbH, 2014, pp. 40, 41.

Appendices

A Data Sets and Reference Values

The Raman spectral data set used for the PCA model is summarized in table A.1, below. 95 spectra are included. The reference values used for the development of the NIR quantitative models are detailed in tables A.2 to A.4), below. All 74 data point used for both PCA and PLS models are included.

Table A.1: Raman spectral data set. In indicates the samples from each accumulation included in this data set.

		Accumulations																
		i	**ii**	**1a**	**2a**	**3a**	**4a**	**5a**	**6a**	**7a**	**8a**	**9a**	**10a**	**11a**	**12a**	**13a**	**14a**	**15a**
Sample number	**0**	In	In	In	In	In	In	-	In	In	In	In	In	In	In	In	In	-
	1	In	In	In	In	In	In	-	In	In	In	In	In	In	In	In	In	-
	2	In	In	In	In	In	In	-	In	In	In	In	-	In	In	-	In	-
	3	In	In	In	In	In	In	-	In	In	In	In	In	In	In	In	In	-
	4	In	In	In	-	In	In	-	In	In	In	In	In	In	In	-	In	-
	5	In	In	In	In	In	In	-	In	In	In	-	In	In	-	In	In	-
	6	-	In	In	-	In	In	-	-	-	-	-	In	-	-	-	-	-
	7	-	In	-	-	-	-	-	-	-	-	-	In	In	-	In	-	-

Table A.2: Reference values for TSS concentration (g/L) obtained through the quantification method detailed in section 3.2. Sample 0 refers to the initial point, before feeding the first pulse, sample 1 refers to the sample collected by the end of the 1st pulse, and so on. Missing data indicates inaccurate or not-available analytical results, meaning the corresponding spectra were not included in the NIR models.

		Accumulation												
		1b	**2b**	**3b**	**4b**	**5b**	**6b**	**7b**	**8b**	**9b**	**10b**	**11b**	**12b**	**13b**
Sample number	**0**	5.27	5.69	6.07	8.25	5.82	6.65	-	6.79	2.5	2.61	3.03	2.06	3.18
	1	5.38	5.04	6.92	8.23	6.59	6.99	6.94	7.35	-	2.74	3.73	2.95	3.40
	2	5.57	5.44	7.47	9.52	8.01	7.25	5.88	7.52	3.45	2.81	4.37	-	3.91
	3	5.85	5.95	8.16	10.22	8.64	6.82	6.21	8.46	3.45	3.32	-	3.11	-
	4	6.09	6.46	8.39	11.05	8.81	7.56	-	8.79	3.58	3.58	-	3.34	3.92
	5	6.25	6.39	8.31	11.12	9.69	-	-	-	-	3.34	6.29	4.16	4.14
	6	6.24	-	-	-	-	-	-	-	-	-	-	3.74	4.12
	7	-	-	-	-	-	-	-	-	-	-	-	3.63	3.72
	8	-	-	-	-	-	-	-	-	-	-	-	-	3.93
	9	-	-	-	-	-	-	-	-	-	-	-	-	4.57

Table A.3: Reference values for PHA content (%) obtained by GC, as detailed in section 3.2. Sample 0 refers to the initial point, before feeding the first pulse, sample 1 refers to the sample collected by the end of the 1st pulse, and so on. Missing data indicates inaccurate or not-available analytical results, meaning the corresponding spectra were not included in the NIR models

		Accumulation												
		1b	**2b**	**3b**	**4b**	**5b**	**6b**	**7b**	**8b**	**9b**	**10b**	**11b**	**12b**	**13b**
	0	6.8	3.2	4.5	6.2	3.0	10.4	-	27.2	7.3	3.8	5.4	4.0	3.7
	1	11.7	18.2	22.2	17.5	17.9	30.7	46.1	42.9	-	27.5	20.3	32.1	24.1
	2	23.7	38.8	31.0	37.0	37.6	45.6	59.9	50.3	36.4	45.0	31.7	-	44.1
	3	34.3	47.8	36.0	52.6	52.7	54.5	55.7	41.7	46.5	47.6	-	27.6	-
Sample	**4**	50.0	48.8	39.5	77.2	60.1	47.7	-	62.0	70.6	63.2	-	48.4	60.0
number	**5**	58.7	50.5	47.4	55.4	52.2	-	-	-	-	45.8	65.3	66.8	64.8
	6	50.0	-	-	-	-	-	-	-	-	-	-	78.8	79.6
	7	-	-	-	-	-	-	-	-	-	-	-	75.1	81.0
	8	-	-	-	-	-	-	-	-	-	-	-	-	84.7
	9	-	-	-	-	-	-	-	-	-	-	-	-	89.0

Table A.4: Reference values for PHA concentration (g/L) obtained from the product of the PHA content (%) and the TSS concentration (g/L). Sample 0 refers to the initial point, before feeding the first pulse, sample 1 refers to the sample collected by the end of the 1st pulse, and so on. Missing data indicates inaccurate or not-available analytical results, meaning the corresponding spectra were not included in the NIR models

		Accumulation												
		1b	**2b**	**3b**	**4b**	**5b**	**6b**	**7b**	**8b**	**9b**	**10b**	**11b**	**12b**	**13b**
	0	0.36	0.18	0.27	0.51	0.17	0.69	-	1.85	0.18	0.10	0.16	0.08	0.12
	1	0.63	0.92	1.54	1.44	1.18	2.15	3.20	3.15	-	0.75	0.76	0.95	0.82
	2	1.32	2.11	2.31	3.52	3.01	3.31	3.52	3.78	1.26	1.27	1.39	-	1.72
	3	2.01	2.84	2.93	5.38	4.55	3.72	3.46	3.53	1.60	1.58	-	0.86	-
Sample	**4**	3.04	3.15	3.31	8.53	5.30	3.61	-	5.45	2.53	2.26	-	1.61	2.35
number	**5**	3.67	3.23	3.94	6.16	5.05	-	-	-	-	1.53	4.11	2.78	2.68
	6	3.12	-	-	-	-	-	-	-	-	-	-	2.95	3.28
	7	-	-	-	-	-	-	-	-	-	-	-	2.73	3.01
	8	-	-	-	-	-	-	-	-	-	-	-	-	3.33
	9	-	-	-	-	-	-	-	-	-	-	-	-	4.07

In the case of PHA content (Table A.3), triplicate measurements were performed through GC and the final PHA content reference value is either the average of the three measurements, the average of two consistent measurements or just one measurement when the other two were clear outliers from the accumulation trend.