New Approaches to Real Time Monitoring of Bio-manufacturing Processes

Scott Paton Parker

Strathclyde Institute of Pharmacy & Biomedical Sciences

A thesis submitted to the University of Strathclyde in part fulfilment of the regulations for the degree of Doctor of Philosophy.

2017

Copyright Declaration

Signed:

Date:

Acknowledgements

First and foremost, I would like to thank my supervisors: Professor Brian McNeil, Professor Linda Harvey and Dr Alison Nordon for their continued support and advice throughout my PhD. I would also like to thank Professor Daniel Bracewell and the other staff members of UCL, for their assistance during this project.

My parents deserve special thanks for their incredible support, encouragement and their constant belief in my abilities throughout my academic career. My brother Ross, Papa Jim and the rest of my family have my thanks for their support. Alison is owed special thanks, for all the patience and understanding she has shown me, as well as her unbelievable ability to keep me positive through the final stages of writing.

I would also like to thank all my friends and colleges within SIPBS, but in particular those in the Strathclyde Fermentation Centre: Dr Mariana Fazenda, Dr Peter Gardner, Dr Chiara Viegelmann, Dr Laura Jeffrey, Dr Wan Bin Mohtar, Joana Faustino, Ellis Robb, Nurzilla Binti Ab Latif and Walter McEwan, for their support throughout my PhD.

Finally I am grateful to the EPSRC Centre for Innovative Manufacturing in Emergent Macromolecular Therapies for providing me with the studentship, and to Lonza Ltd and Thermo Fisher Scientific for providing me with all the materials to carry out this research.

Table of Contents

List of Figures & Tables

Abbreviations

ATR	Attenuated Total Reflection
CD	Chemically Defined
CHO	Chinese Hamster Ovary
CO_2	Carbon Dioxide
DMSO	Dimethyl Sulfoxide
dO_2	Dissolved Oxygen
DoE	Design of Experiment
EFA	Efficient Feed A
EFB	Efficient Feed B
ELISA	Enzyme-Linked Immunosorbent Assay
FT	Fourier Transformed
FT-IR	Fourier Transform Infrared
H_2O_2	Hydrogen Peroxide
hIgG	Human Immunoglobulin G
HPLC	High-Performance Liquid Chromatography
IR	Infrared
LoD	Limits of Detection
mAb	Monoclonal Antibody
MIR	Middle Infrared
miRNA	Micro Ribonucleic Acid
MLR	Multiple Linear Regression
MSC	Multiplicative Scatter Correction
MSX	Methionine Sulfoximine
NIR	Near Infrared
PAT	Process Analytical Technology
PC	Principal Component
PCA	Principal Component Analysis

PLS	Partial Least-Squares
QbD	Quality by Design
RMSE	Root Mean Square Error
RMSEC	Root Mean Square Error of Calibration
RMSEV	Root Mean Square Error of Validation
RNA	Ribonucleic Acid
SG	Savitzky–Golay
SNV	Standard Normal Variate
v/v	Volume/Volume Percent
ZnSe	Zinc Selenide

Abstract

Over the past decade the bioprocessing industry has embraced and led the creation of a number of initiatives to drive process development and optimisation. There has consequently been an increased demand for tools capable of meeting the requirements of these initiatives. Spectroscopic monitoring systems such as Near infrared (NIR), Mid infrared (MIR) and Raman can provide a number of substantial advantages over more conventional off-line monitoring methods, normally applied to bioprocessing. With these techniques the bioprocessing industry has means of meeting these demanding challenges.

The aim of study was to determine the feasibility of using NIR, MIR and Raman spectroscopy as a combined 'toolkit' to correlate changes in metabolic profile to spectral changes, within low passage Chinese Hamster Ovary cell cultures. The target metabolites focused upon in the spectroscopic analysis for this study were glucose and lactate. Conventional off-line techniques were used to investigate any correlation between spectral changes and those in the reference data. A design of experiment (DoE) approach was used to identify the optimum preprocessing techniques of the data and to build more accurate process trajectory models. These models were subjected to both an internal and external validation to guarantee the reliability of the results. Feasibility of data fusion, to create a single 'fused' dataset from the three spectroscopic techniques was also assessed, to produce a model with lower errors of prediction that the individual parts. The other area investigated was the characterisation of low passage number cultures in various media and process formats, which was of interest to the industrial collaborators. To test this three media; CD CHO, CD OptiCHO and Dynamis were provided by Thermo Fisher Scientific. Currently there is no system which utilises this toolkit with a combined DoE and data fusion

strategy, within the bioprocessing industry. This research demonstrates a gap in the industry and a novel approach as to how tackle process monitoring.

The results in this project demonstrate the varying degrees of success the spectroscopic techniques across all process formats and media. The NIR proved to be the most successful at modelling the target metabolites, with MIR being unsuccessful and Raman only being able to detect but not model the metabolites. This research provides an indicator of media suitability in low passage number cell cultures, by comparing the batch culture processes. The CD CHO media proved to be the best of the three tested media, based upon the cell density, viability and mAb titre produced. Overall this study represents a valuable stage in the progression towards real time monitoring of biomanufacturing processes and development of tailored low passage number cell culture media. However there are areas of this study where further investigation could be improved.

Chapter 1:

Introduction

1.1 Introduction

As a result of the bioprocessing industry embracing the PAT and QbD initiatives, which seek to apply new technologies and build quality into processes respectively, there is a need to implement tools capable of meeting these challenges and moving towards real time bioprocess monitoring. Utilising spectroscopic technologies such as Near infrared (NIR), Mid infrared (MIR) and Raman offers industry a means of meeting these needs. The overlapping nature of the data produced by these three technologies is what could potentially offer the opportunity to gain an improved process understanding and increased process control. Consequently this combination could lead to improved final product quality and potentially lead to greater yields. However, these techniques haven't been investigated as a combined 'toolkit' for the purposes of bioprocess monitoring.

In the biopharmaceutical industry there are a number of expression systems used to produce therapeutic agents, of these, mammalian cell lines are often utilised due to their ability to produce more 'humanised' therapeutic proteins as they have reduced adverse effects in the clinic. One category of therapeutic proteins, monoclonal antibodies represent a large section of the market due to their impact on both clinical and medical research. As such, production of therapeutic proteins requires constant monitoring to ensure that no deviations occur within the production process. This again is where spectroscopic techniques offer the opportunity to provide constant monitoring.

This study utilised a high yielding monoclonal antibody (hIgG) producing industrial Chinese Hamster Ovary (CHO) cell line, due to their extensive use within the biopharmaceutical industry. This cell line was provided by a world leading supplier to the pharmaceutical and biotechnology sectors, Lonza, who were interested in collaborating in this research. Their interests coincided with the other industrial collaborator involved,

Thermo Fisher Scientific, who provided all the culture media and feeds. The interests of both companies overlapped well as both had an interest in characterisation of the cell line in various media and process formats. However Thermo Fisher had a specific interest in investigating low passage number, as it would help them better develop media to aid in early cell culture development. As such the needs of the collaborators were integrated into this research, with a low passage number culture grown in increasingly more complex medium and format (Table 1.1).

Table 1.1 Experimental process methodology

Reactor Passage Number	Medium	Process Format
7	CD CHO	Batch
		Fed Batch (EFB)
	CD OptiCHO	Batch
		Fed Batch (EFA)
	Dynamis	Batch

The objective of this research was to move towards a real time monitoring system and development of process models, through the application of an at-line monitoring system incorporating NIR, MIR and Raman spectroscopies. An at-line system is where the analyser is not directly implemented within a bioprocess system, but is within the bioprocessing area. While an on-line monitoring system would be more suitable in the development of real time process analysis, they are far more complex in terms of implementation/integration. Therefore an at-line monitoring system was more suitable for establishing a baseline monitoring system, during the investigation. The other novel aspect of this research was the design of experiment approach that was taken in the preprocessing of the spectral data. The aim of investigating this methodology was to improve the final process models, by taking a systematic approach to preprocessing combined with prior

knowledge. While this approach has been demonstrated to have some merit in chemical processing, it has yet to be applied to a biopharmaceutical process. The final novel feature to be investigated was determining the feasibility of applying data fusion, with the aim of combining the individual datasets from each technique, if suitable. The result of this would be a single 'fused' dataset which when modelled should produce a process model with reduced prediction errors than would be present from the three individual models.

Throughout this study, application of NIR, MIR and Raman to the various cell culture processes for the development of process models is documented and unbiasedly presented. Both the successes and the limitations of the applied techniques are assessed and discussed. From these assessments, conclusions are presented alongside the potential future developments for the research in this study.

Chapter 2:

Literature Review

2.1 Bioprocess monitoring

2.1.1 Why monitor bioprocesses?

Over the past few decades increasing importance and interest has been placed on bioprocess monitoring. Effective process monitoring has become essential in academia, as a means to study microbial physiology and in industrial bioprocess development where it provides a time-efficient method of extracting more precise data from each experiment (Olsson *et al.*, 1998, Lourenco *et al.*, 2012). Within many industrial sectors such as beverage production, pharmaceutical, agro-food and biowaste treatment, where demand to have decreased process times and cost effective means of product analysis is required, bioprocess monitoring has steadily been more widely implemented. Such implementation in some industries leads to reduced costs, increased productivity and efficiency, improved process reproducibility and quality control (Vojinovic *et al.*, 2006, Ündey *et al.*, 2010, Svendsen *et al.*, 2015).

Different industrial sectors, despite the varying prices of their end products and their goals, are connected in their requirement for cultivation of microorganisms, mammalian cells, insect cells and plant cells, processes which require accurate monitoring systems. To cultivate cells in bioreactors is challenging with regards to control and monitoring, since during the running of the process, substrates are utilised and intermediates and products are produced in the presence of living cells (Workman *et al.*, 2003). Small uncontrolled changes in nutrient availability, temperature, pH, and pressure can have significant effects on cellular metabolism and can alter the overall efficiency and productivity, and can even result in the process being unprofitable (Sales *et al.*, 2015). Full monitoring and control of bioprocesses aims to control the environment within the reactor

to the optimal conditions for growth, biosynthesis, and/or downstream processing (Vojinovic *et al.*, 2006).

It should also be recognised that while intensive bioprocess monitoring is a must in the biopharmaceutical industry, and can theoretically be applied to any industrial process, in reality it is often not economically viable for use in all types of industrial bioprocesses. This is particularly true of bioprocesses where the end product either does not require such strict regulations as biopharmaceuticals or is not of high value, such as Baker's yeast. In the biopharmaceutical industry, there is pressure to develop more cost-effective manufacturing processes, and innovating the way in which bioprocesses are monitored offers one solution to this problem. However, the development and application of bioprocess monitoring can make up a significant portion of the overall investment made in bringing a product to market (Harms *et al.*, 2002). It has previously been outlined that the costs associated with process development and clinical manufacturing, which bioprocess monitoring falls under, can represent up to 40-60% of the overall investment. Thus the cost of purchasing and implementation of monitoring equipment must be heavily assessed and scrutinised (Rosenberg, 2000, Farid, 2007).

Despite the need for better process monitoring, there is still a significant way to go, especially in the biopharmaceutical industry (Zhao *et al.*, 2015). One study in particular summaries this disparity very well by stating, while there have been advances in academic fields the translation of techniques into industrial settings is significantly lacking, noting that "most" industrial processes still employ "primitive" strategies. Again the authors demonstrate that one of the main reasons for this is due to the high cost of new technologies (Mercier *et al.*, 2013). Regardless of the need for up-front investment, if bioprocess monitoring development is carried out it could result not only in reduced

production costs and improved product yields, but also an overall decrease in future development costs (Clementschitsch and Bayer, 2006, Mercier *et al.*, 2013).

However in recent years there have been significant steps and initiatives taken by both regulators and the biopharmaceutical industry, to address the shortcomings in the development of bioprocesses and their monitoring (Gomes *et al.*, 2015).

2.1.2 Current Trends in Biopharmaceutical Monitoring

2.1.2.1 Process Analytical Technology

Biopharmaceuticals have a stringent regulatory requirement, since they are products for use on/in humans and thus need to be manufactured by means that provide the highest possible quality assurance (Miller *et al.*, 1996). To meet the need for high quality assurance, the United States Federal Drug Administration released a series of voluntary guidelines entitled, "Process Analytical Technology - A framework for innovative pharmaceutical development, manufacturing, and quality assurance" (F.D.A., 2004). The document outlined a series of voluntary guidelines which would lead to the development and implementation of innovative pharmaceutical product development, manufacturing, and quality assurance.

Implementation of PAT within the biopharmaceutical industry is essential for a number of reasons. The primary reason is that the consumer's health is paramount and so the more data obtained regarding the quality of a biopharmaceutical, the greater the confidence there is when it is released to the clinic. Another reason is the need for companies to prove to regulators that the highest possible standards of production are being met. These are two of the reasons that PAT has been implemented within the biopharmaceutical industry, but what is PAT? (Swarbrick, 2007).

In general terms, PAT can be seen as any technology that integrates aspects of microbiological, chemical, mathematical and physical analysis in a given bioprocess. Implementation of PAT aims to result in the highest possible quality of product, through thorough monitoring and control of the bioprocess (F.D.A., 2004, Gomes *et al.*, 2015, Murphy *et al.*, 2016). The technologies utilised in PAT cover a range of different functions, e.g. multivariate tools for the purpose of design, acquisition of data, and analysis or process analysers and control tools (De Beer *et al.*, 2008). Figure 2.1, shows the three key steps by which such tools can be used, often in combination, to achieve the PAT goals of a particular bioprocess.

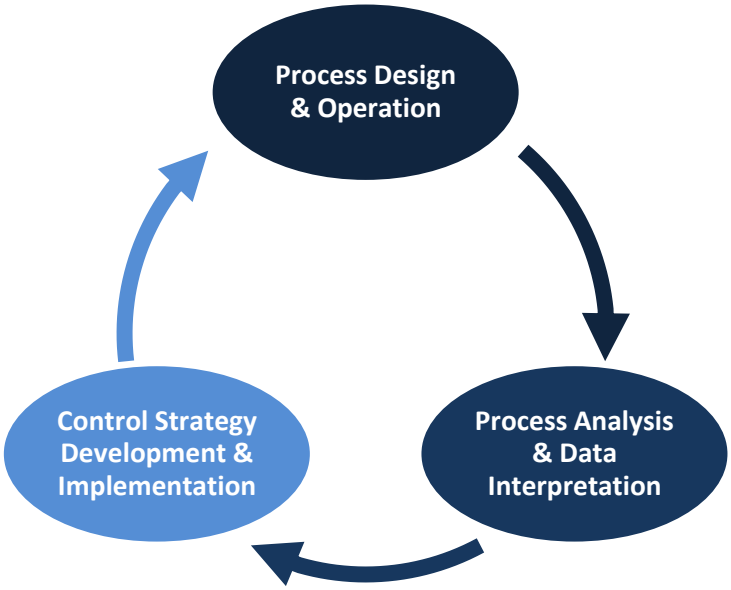


Figure 2.1: PAT Quality Assurance Steps. Adapted from; (Read *et al.*, 2010).

In a PAT implemented bioprocess there are three methods which utilise real-time or near real-time bioprocess measurements; firstly at-line, or near real-time, involving manual sampling which is then transported to an analyser located in the processing area; secondly, on-line, where sample extraction and transport through a sample line is automated. The sample in the line is subsequently sent to an automated analyser and once

analysed the sample is returned to the process stream via another transport line; and finally *in-situ*, this requires no removal of samples as a probe connected to the analyser can be placed inside the reactor vessel. The latter two methods are the most desirable as they provide direct analysis without destruction of or interference with the sample (F.D.A., 2004, Murphy *et al.*, 2016).

In order to produce the highest possible quality of product, coupling PAT with risk assessment strategies is required to provide measurement of critical quality attributes. These attributes result in alterations to the bioprocess in order to maintain the high quality of the product and the process is subsequently validated. The only way to effectively improve product quality is by ensuring bioprocess reproducibility which can be done by PAT implementation. Through use of the data obtained, where real-time trends within the bioprocess can be monitored and controlled PAT has the capability to greatly improve the production of biopharmaceuticals (Challa and Potumarthi, 2012, Gomes *et al.*, 2015).

2.1.2.2 Quality by Design

Since the development of the PAT initiative in 2004, further work to regulate and standardise the production of biopharmaceuticals has been carried out, so the Quality by Design (QbD) initiative has been introduced. QbD encompasses a number of overlapping concepts which include PAT, risk- assessment and management, creation of a bioprocess knowledge database and process design spaces. The ICH pharmaceutical development guideline Q8 (R2) defined the objective of QbD in 2009, as an approach to build a high degree of end product quality into the bioprocess, instead of just testing for it post-process (Murphy *et al.*, 2016). At the heart of QbD is the intention to understand the relationship between the overall efficacy and safety of the end product, quality of the product, and the means by which the product is manufactured (Read *et al.*, 2010, Rouiller *et al.*, 2012,

Gomes *et al.*, 2015, Junker *et al.*, 2015, Murphy *et al.*, 2016). The result of understanding such factors means that companies can constantly assess and update their bioprocesses when needed, to guarantee a high level of consistency in their end product. The QbD approach to bioprocessing of biopharmaceuticals is very different from the more traditional approaches, which rely on repeated testing to ensure product quality (Rathore and Winkle, 2009, Rathore, 2016) (Figure 2.2).

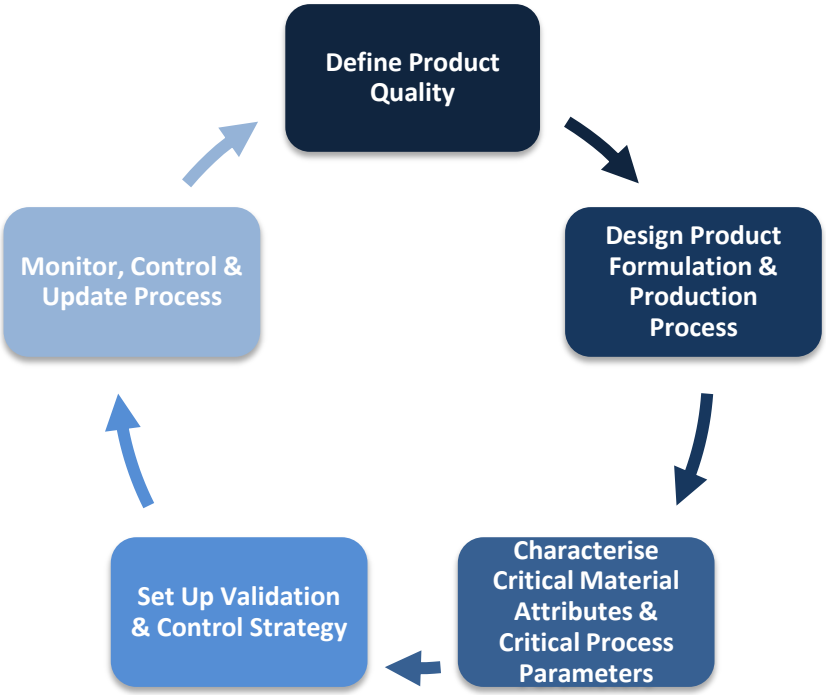


Figure 2.2: Overview of QbD Strategy. Adapted from; (Rathore and Winkle, 2009).

Application of QbD seeks to achieve manufacture of a product with a pre-defined quality and so when implementing QbD the first stage is to gather information. The gathering of information must be extensive and must include; the product's therapeutic action, initial raw materials used for production, means of product manufacturing and all possible sources of variability. Once this information is gathered, it is used to set the bioprocess parameters and coupled with continuous real-time monitoring, helps ensure

that the end product meets not only the pre-defined quality but also the highest level of safety and efficacy (Rathore and Winkle, 2009, Gomes *et al.*, 2015, Rathore, 2016).

Since the QbD approach was first introduced, the biopharmaceutical industry has been able to implement its framework into the production of small-molecule therapeutics. However in contrast, implementation of QbD into the production of protein therapeutics has been more challenging. The most probable reason for this is that the manufacturing of small-molecule therapeutics is better understood, and thus, the relationship between the variable aspects of the process and the end product quality is much easier to define. In the production of protein therapeutics there are a number of factors which could explain why application of QbD is difficult such as, the complex structures and the need for purification of the target protein. However, although there have been difficulties in implementing QbD in this area, there is a belief among regulators and researchers that there is enough knowledge either currently available or that can be obtained through experimentation and modelling, which would elucidate the relationship between the bioprocess and the protein therapeutic quality, that allows for successful implementation of QbD (Rathore and Winkle, 2009, Del Val *et al.*, 2012, Junker *et al.*, 2015).

Implementation of QbD in the biopharmaceutical industry is believed to result in reduced process approval costs and time, lower regulator interference, and potentially lead to better optimisation and innovation, as bioprocesses can be developed based on the relationship between the variable aspects of the process and the end product quality (Kourti and Davis, 2012, Kumar and Gupta, 2015). As this relationship is defined by the understanding and application of scientific and engineering principles, the results of the process are easier to predict, and thus would decrease regulatory approval times and the cost of process development. The predictions would also lead to more efficient controls

systems, which would greatly reduce the possibility of the end product failing to meet an acceptable standard of quality. The information gathered during QbD implementation and the resulting lessening in regulatory approval steps would greatly contribute to optimisation and potential development of novel methods processing (Yu, 2008, Rathore and Winkle, 2009, Del Val *et al.*, 2012, Rathore, 2016).

2.1.3 Addressing PAT/QbD

As a result of the drive to embrace and implement PAT and QbD to innovate bioprocess monitoring, there has been increased need to implement tools suitable for meeting the requirements of these initiatives. Tools that would meet the requirements of PAT and QbD for bioprocess monitoring would need to be capable of continuous real time monitoring of the variable factors, substrates, intermediates and products, within a given process. Given these needs, spectroscopic sensors offer a viable option for meeting the specific demands of both of these initiatives and provide a means for innovating bioprocess monitoring.

2.2 Spectroscopic monitoring

Spectroscopic sensors such as Near infrared (NIR), Mid infrared (MIR) (Figure 2.3) and Raman provide a number of substantial advantages in the monitoring and control of bioprocess. They can be integrated in-line and thus there is no need to remove samples from the process. These sensors can be miniaturized and multiplexed, and can also send data instantaneously. The most beneficial advantage of these technologies is that they can potentially monitor several analytes simultaneously and be used together to create a more comprehensive overview of the bioprocess, allowing for greater monitoring and control. However, in order to obtain the desired data, these techniques must be used in conjunction

with multivariate analytical techniques like, principal component analysis (PCA) and partial least squares analysis (PLA) (Landgrebe *et al.*, 2010, Biechele *et al.*, 2015, Glassey, 2016).

Figure 2.3: Types of radiation and their approximate wavelength ranges within the electromagnetic spectrum. Adapted from; (Lourenco *et al.*, 2012).

Spectroscopic sample analysis can be broken down into three main categories; off-line, at-line, which is dependent on the location of the sensor to the bioreactor, and on-line (Figure 2.4) (Lourenco *et al.*, 2012, Sales *et al.*, 2015).

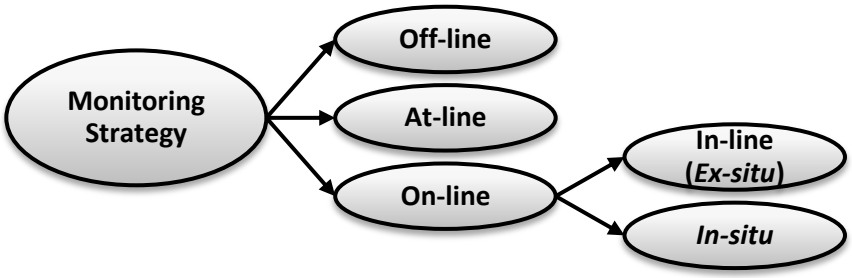


Figure 2.4: Spectroscopic sampling approaches. Adapted from; (Lourenco *et al.*, 2012)

Off-line sampling is a destructive process, in that requires the removal of samples from the reactor for analysis. These samples are often then stored or pre-treated in some way before being transferred, to a separate analysis laboratory (Vaidyanathan *et al.*, 1999). The main downside to this approach is the time period between removing a sample, until the analysis is completed. In regards to process control this lag period is not desirable, as the opportunity to correct any process deviations will have passed by the time the data

from the analysis has been gathered and the possibility of significant change within the removed sample (Sales *et al.*, 2015).

At-line sampling has a similar methodology to that used for off-line analysis. It also requires either manual or automatic removal of a sample from the bioreactor, but instead of transporting the removed sample to a separate analysis laboratory it is analysed using instrumentation nearby the bioreactor. At-line sampling offers a slightly faster analysis option than the off-line method, however the process is still destructive and has an undesirable time lag between sampling and analysis (Lourenco *et al.*, 2012, Sales *et al.*, 2015).

The final method, on-line, unlike the previous two methods, integrates sampling analysis into the process allowing near real time acquisition of process data. On-line can also be subdivided into either a, *ex-situ* (in-line) or *in-situ* configuration (Vaidyanathan *et al.*, 1999, Lourenco *et al.*, 2012, Sales *et al.*, 2015) (Figure 2.5).

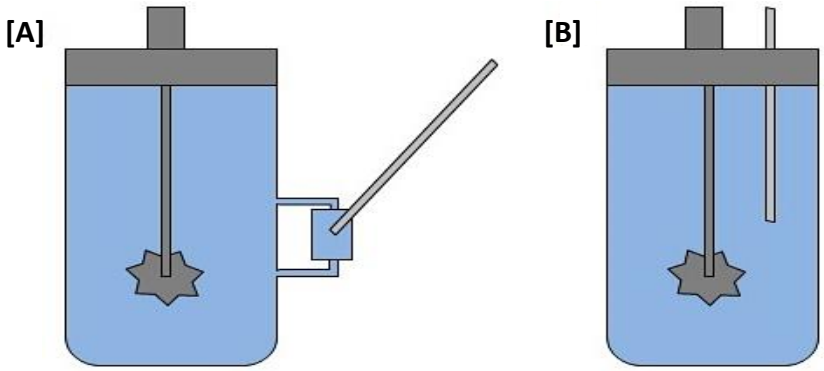


Figure 2.5: On-line sample analysis – [A] *Ex-situ* (in-line) probe [B] *In-situ* probe. Adapted from; (Lourenco *et al.*, 2012)

Ex-situ, also called in-line, has the analysis probe outside of the bioreactor vessel with a constant flow of the culture diverted through a flow cell for measurement. An *in-situ*

approach utilises a probe or sensor, inside the vessel itself, coming into direct contact with the culture. This strategy is regarded as the optimum method, as it allows for real time process analysis and does not require any removal of samples for measurement thus avoiding the risk of introducing artefacts in the sample (Vaidyanathan *et al.*, 1999, Lourenco *et al.*, 2012).

2.2.1 NIR Spectroscopy

The region of the electromagnetic spectrum that corresponds to near infrared is located between 780–2500 nm (Figure 2.3) and within this range a number of the functional groups of molecules have had their vibrational frequencies characterised. These molecular vibrations are characterised as a variation in either the length of the bond present, called stretching, or a change in the angle of the bond, called bending. The stretching or bending of the bond can either be symmetrical, which is IR inactive, or asymmetric, which is IR active. For absorption to take place the asymmetric stretching and bending has to lead to a change in the molecule's polarity, a dipole moment. This change and subsequent absorption are dependent on three main factors, the number, size and orientation of atoms (Workman and Weyer, 2012). Due to a high dipole moment of the vibrations arising from covalent bonds, and the presence of such bonds in nearly every biological molecule, they form the foundations of the data acquired from NIR and thus NIR can be used in the monitoring of bioprocesses (Scarff *et al.*, 2006).

Implementation of NIR utilising *in situ* fibre-optic probes has been shown to produce excellent results when used for the characterising and optimisation of process conditions, and in the monitoring of substrates, intermediates and products, and elucidation of the mechanisms for multicomponent mixtures (Mcfearin *et al.*, 2011, Zhao *et al.*, 2015). These NIR probes offer a few advantages, one example is the ability to provide

comprehensive non-destructive measurements as the probe is *in situ*, so no samples needs to be removed from the process stream (Genkawa *et al.*, 2012). Another advantage provided by fibre-optics is that they exhibit high light transmission and can hence be much longer in length than other materials and this advantage allows for the spectrometer and the light source to be located away from the processing area. The use of fibre-optics also allows for multiplexing of probes to different process reactors connected to a single spectrometer, thus eliminating the need for multiple systems (Roychoudhury *et al.*, 2007, Beuermann *et al.*, 2012).

NIR is a useful technique for the gathering of real time reaction data, for enabling a high degree of product quality and reproducibility, providing modelling data and overall control of the process thus contributing to process optimization (Becker *et al.*, 2007). As NIR is a relatively insensitive technique with weaker absorbance than that of MIR, 10-100 times weaker, thus it enables samples to be directly analysed (Scarff *et al.*, 2006, Tamburini *et al.*, 2014). Initial application of NIRS for bioprocess monitoring was carried out by Karl Norris in the 1980's and he applied it in solid-state fermentations for quantitative analysis (Hakemeyer *et al.*, 2012). Since then NIRS has been thoroughly applied in many industries as means of process monitoring and quality assurance for initial materials, intermediates and end product testing (Scarff *et al.*, 2006, Vanarase *et al.*, 2010, Sinelli *et al.*, 2011, Grassi *et al.*, 2014, Glassey, 2016, Murphy *et al.*, 2016).

While the limits of detection for NIR may be in the g/L range, this can be further lowered to approximately 100 ppm by investigating the effects target components have on the NIR spectra, through the use of multivariate data analysis techniques. (Hakemeyer *et al.*, 2012, Tamburini *et al.*, 2014). As multivariate data analysis on the NIR spectrum bases the predictions using all of the process factors this allows for, more accurate process

monitoring, predictions on batch quality, fault detection and hence process optimisation (Hakemeyer *et al.*, 2012). Some of the most common pre-processing methods used to analyse the data include, centering of data, baseline correction and the scaling of the data in order to improve the predictability of the evaluation method, to name a few (Beutel and Henkel, 2011). This analysis is essential in industry as the benefits previously mentioned can lead to reduction in delays and allow any preventive measures to be taken far more quickly than with other monitoring techniques, thus make the process more cost-effective (Scarff *et al.*, 2006).

2.2.2 MIR Spectroscopy

Unlike the NIR region which is characterised by only overtones and vibration combinations, the mid infrared region, 2500–25000 nm (Figure 2.3), is characterised by excitations of fundamental molecular vibrations. The spectra obtained from MIR also differ from NIR, in that it produces well defined peaks for the scanned material (Landgrebe *et al.*, 2010, Sales *et al.*, 2015).

Within the MIR spectra there are four regions which each correspond to specific stretches (Table 2.1). Region 1 is associated with fatty acids and within Region 2 there are a few stretching bands, the first of which is the lipid based C=O stretching band. The next bands are those which are associated with proteins and peptides. These bands are important for interpreting the structural components of the molecule based on the α-helix, β-sheet and random coil conformations. Region 3, corresponds to a number of cellular components, such as phospholipids and nucleic acids and finally, Region 4 of the spectra relates to the structure of polysaccharides (Lu *et al.*, 2011).

Table 2.1: MIR regions and their associated structural features.

Region of the Spectrum	1	2	3	4
Spectral Location (cm^{-1})	3,000–2,800	1,800–1,500 (specific bands) [i - 1,740] [ii - 1,650] [iii - 1,550]	1,500 and 1,200	1,200 and 900
Structural Feature	aliphatic C—H stretches	i – lipid based C=O stretching ii – amide I iii – amide II	phospholipids and nucleic acids	polysaccharides

Implementation of MIR for the quantification of multiple components in aqueous solutions has advanced greatly in the past decade to become an on-line technology. This can in part be related to attenuated total reflection (ATR) sample presentation mode and the application of the mathematical technique, Fourier transformation (FT). Fourier transformation minimises the issues that are associated with signal stability and measurement noise. FT-MIR which utilise interferometers as opposed to monochromators, is more beneficial for quantification of concentrations of components in complex mixtures, as opposed to NIR. Overall the application of FT-MIR has made it easier for spectra acquisition of multiple compounds in complex aqueous mixtures (Roychoudhury *et al.*, 2006b, Schenk *et al.*, 2007).

There are however some limitations of MIR, firstly as infrared absorption requires a change in dipole moment to occur and certain polar groups have strong stretching vibrations that can interfere with the spectra. One example of a polar molecule that has a strong infrared absorption is water and thus when analysing the spectra this interference needs to be compensated for. Another limitation of MIR is the limit of detection is potentially not low enough to allow for quantification of low levels of analytes in contrast

to techniques like immunochemical analysis (Lu *et al.*, 2011, Foley *et al.*, 2012). Finally as with NIR, one method to quantitatively analyse the data is by the application of multivariate data analysis, such as Partial Least-Squares (PLS). MIR is potentially capable of producing data on large range of products and offers an excellent method for the qualitative analysis of compounds (Schenk *et al.*, 2007, Capito *et al.*, 2013).

The advantages that MIR offers, high absorption capacity and well defined peaks, are what make it an invaluable technique for the monitoring of biological processes (Capito *et al.*, 2015b). But the development and application of FT-MIR in combination with multivariate data analysis techniques to produce process models, allowing identification and quantification of different analytes, is why MIR has increasingly been applied as a bioprocessing monitoring technique (Roychoudhury *et al.*, 2006b, Capito *et al.*, 2013).

2.2.3 Raman Spectroscopy

Raman spectroscopy utilises the scattering of photons as opposed to NIR and MIR which rely on absorbance of light (Paudel *et al.*, 2015). When a photon of light interacts with a molecule, it can result in the induction of a transition in the energy states, which causes the molecule to be in an excited vibrational state. The majority of photons that are scattered by the molecule have the same energy as the initial light ray that strikes it and this is referred to as Rayleigh (elastic) scattering. However a minority of the scattered photons possess a different frequency and this is referred to as Raman (inelastic) scattering (Ashton and Goodacre, 2011, Zhao *et al.*, 2015). As Raman scattering is dependent on a change in the dipole moment of functional groups during molecular vibrations, S—S and C—C bonds which have nonpolar groups have strong Raman bands (Kneipp and Kneipp, 2006, Lu *et al.*, 2011). Depending on the frequency of the Raman scattered photon it can be classed as either Stokes scattering, where it is less than the incident frequency, or anti-Stokes

scattering, where it is greater than the incident frequency. Stokes scattering is the most common form measured and it provides greater Raman intensities (Ashton and Goodacre, 2011) (Figure 2.6).

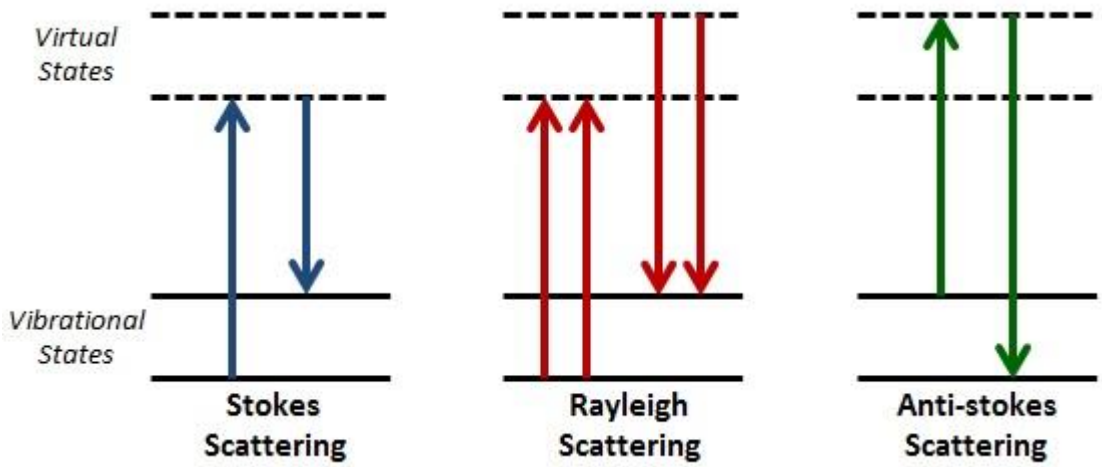


Figure 2.6: Energy level diagram of Raman and Rayleigh scattering. Adapted from; (Ashton and Goodacre, 2011).

Raman offers a number of advantages over the infrared spectroscopy techniques for the monitoring of biological processes, as the interference of water on the spectra is reduced. Over the same wavenumber range, Raman produces more detectable spectral features than that of infrared. Bands produced from Raman have a tendency to be narrower than those in the MIR range and this is due to the increased specificity offered by Raman lasers/detectors. The potential analytical wavelength range and laser resources are also greater for Raman than for the infrared technologies (Lu *et al.*, 2011, Zhao *et al.*, 2015). Another property that makes Raman a useful probe in the monitoring of bioprocesses is that there is comprehensive structural information produced within the spectra. When monitoring proteins in an aqueous medium, the orientation of hydrogen bonds and the interactions of side chains can be determined by monitoring the process environment and this information can subsequently be used to determine the stability of the protein. Further

combining the side chain stability data with the conformational details of secondary structure elements, makes Raman an exceptional tool for monitoring of biological systems (Ashton and Goodacre, 2011). Finally, bioprocess monitoring using Raman, as with NIR and MIR, also offers the advantage of no sample preparation being required and can be applied to aqueous systems with ease (Abu-Absi *et al.*, 2011).

The main obstacle to the application of Raman spectroscopy in monitoring bioprocesses is the fluorescence produced by many biological molecules which can obscure the Raman spectra. Even if the fluorescent intensity is not great, Raman spectrometers are designed to detect weak signals and the resulting effect of this detection is that Raman bands will seem like small, narrow peaks (Murphy *et al.*, 2016). However there are a number of solutions to this issue, one solution although not necessarily viable is to remove the fluorescent compounds from the sample, however if the fluorescence is from the compound of interest this is obviously not viable (Beutel and Henkel, 2011). A more viable solution is to shift the excitation wavelength to 785 nm or greater, although this reduces the fluorescence it can decrease the efficiency of the scattering in contrast to the 532-633 nm range. The most viable and beneficial solution is to use wavelengths in the deep UV, 180-260 nm, as excitation in this range completely removes fluorescence and can also increase the scattering signal by a factor of 10^3-10^5, compared to normal Raman scattering (Ashton and Goodacre, 2011, Paudel *et al.*, 2015).

The continuing developments in miniaturisation of equipment, reduction in costs and improving software capabilities, are all contributing to the advance of Raman spectroscopy. Overall due to these advances, Raman has been used extensively for a wide variety of applications and appears to be the most promising spectroscopic method for in-

line analysis of complex cell culture systems (Ashton and Goodacre, 2011, Abu-Absi *et al.*, 2011, Singh *et al.*, 2015).

2.3 Chemometrics and Data Analysis

Chemometrics is the application of mathematical and statistical methods to extract and analyse information derived from a combination of reference and spectral data (Murphy *et al.*, 2016). The spectroscopic data collected by NIR, MIR and Raman can contain information about numerous components from within the sample matrix. The effects and correlations of these various components on the resulting spectra generally render univariate analysis techniques unsuitable, thus multivariate analysis methods must be used (Rathore *et al.*, 2011, Rathore and Singh, 2015).

2.3.1 Principal Component Analysis

One method of multivariate data analysis that can be used to reduce the complex spectral data collected from NIR and MIR is principal component analysis (PCA). The basic concept of PCA is to describe the data collected using only abstract variables. These variables are known as Principal Components (PCs), which are representative of those features which have the biggest influence on the process (Abdi and Williams, 2010).

The most basic means of describing the points of a dataset is by expressing them as two-dimensional X and Y co-ordinates. However PCA can be used to further reduce the dimensionality into a single axis with each point represented as a linear combination of the original variables. Principal component one (PC1) is inserted through the data, which describes the maximum variance observed within the samples, and the initial points from the data are projected onto PC1. All subsequent PCs, e.g. PC2, PC3, etc., are positioned orthogonal to the previously plotted PCs and describe increasingly less observed variance in

the dataset. The result of using PCA is that the total variance present in the data, both substantial and random noise, is displayed as a series of PCs and the number of which normally equals the number of process variables (Figure 2.7). PCs representative of random noise can be disregarded, thus reducing the dataset further to include only those PCs that describe significant variation in the data (Miller and Miller, 2000, Gemperline, 2006a).

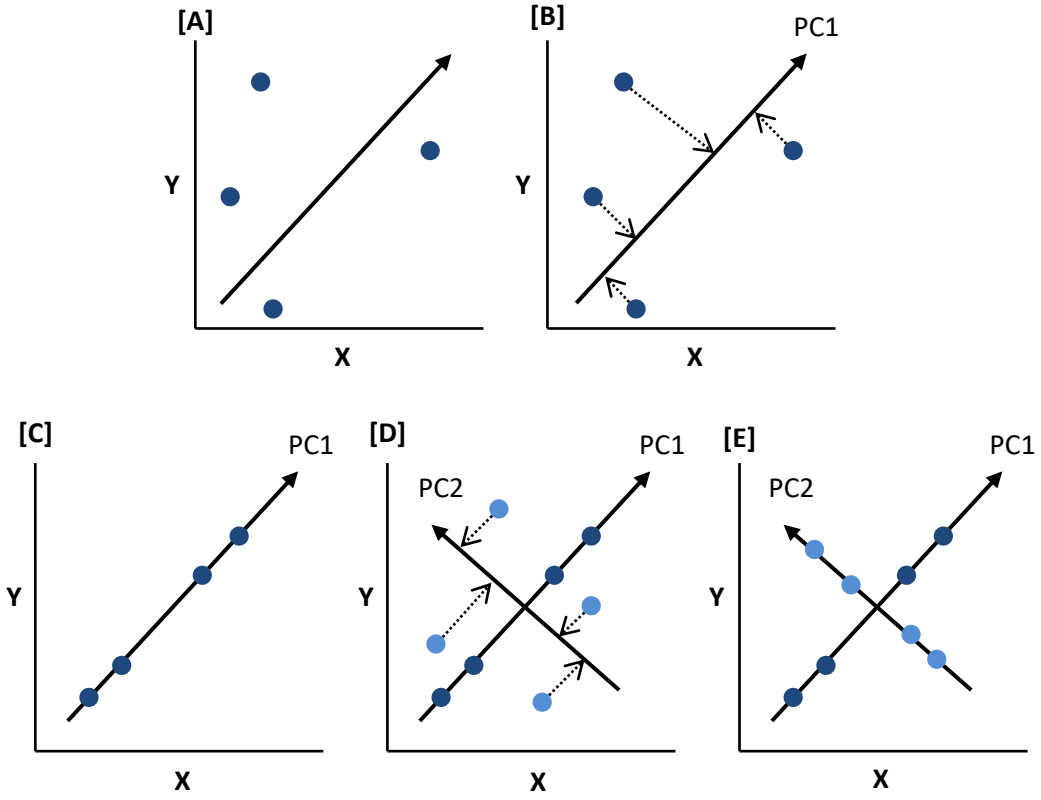


Figure 2.7: Diagrammatic Representation of PCA – [A] data points [B] PC1 inserted showing maximum variance with data point projections [C] data points projected onto PC1 [D] PC2 plotted orthogonally to PC1 with initial data point projections [E] data points projected onto PC2.

PCA generates two small matrices and an error matrix from the original data matrix. One matrix is the scores and this provides information about the samples and how

they relate to one another. The other matrix, the loadings, is representative of how the measured variables are related. In regards to spectral data the scores relate to the sample's spectra and the loadings are representative of the spectral regions causing the variation. Each unique source of variance within the original data matrix is characterised by the total of its constituent principal components, which have their own scores and loadings matrices (Gemperline, 2006a, Abdi and Williams, 2010).

2.3.2 Spectral Pre-processing

The application of pre-processing spectral data is an essential component of chemometrics modelling. The intention of pre-processing is to remove physical phenomena in the spectra thus enhancing the quality of the data and subsequently improve the final model thus improving the accuracy of the predicted analyte concentrations. Two of the most widely used pre-processing techniques are, mean centering of the data and Savitzky-Golay derivation (Miller and Miller, 2000, Gemperline, 2006a, Rinnan *et al.*, 2009).

2.3.2.1 Mean Centering

Mean centering data works by firstly using the average of the total values for each column, which in the case of spectral data represents a wavelength. Then this value is subtracted from the value of each row of that column, which is representative of the spectrum of a sample. The points within the data matrix therefore have no associated mean value, thus dataset origin point is shifted (Wise and Kowalski, 1995, Gemperline, 2006a, Varmuza and Filzmoser, 2016).

2.3.2.2 Savitzky-Golay Smoothing & Derivation

The Savitzky-Golay smoothing and derivation is a polynomial smoothing technique, which uses both a smoothing step and derivation on the spectral data. The aim of the smoothing

step is to enhance the signal while minimising the effects generated by background noise. Spectral smoothing is carried out firstly by taking a single point in the spectrum and examining a region surrounding it. From this examination a mean value is calculated for the region and substituted in place of the original value. This process is then is then repeated for each point across the spectral dataset (Zhang *et al.*, 2005, Gemperline, 2006a, Varmuza and Filzmoser, 2016).

Application of derivatising the spectral data seeks to enhance spectral features and reduce any baseline errors, through narrowing and sharpening the spectral peaks. Similar to the smoothing step, derivation places a narrow window around the data and fits a polynomial expression within it. Coefficients from the polynomial expression are used to calculate the derivative of the given data point. Like the smoothing step, once a point has been determined the window is moved forward repeated across the data (Wise and Kowalski, 1995, Gemperline, 2006a).

2.3.2.3 Multiplicative Scatter Correction (MSC)

Multiplicative scatter correction (MSC) is a spectral pre-processing technique that was initially characterised and then developed in the 1980s by; (Martens *et al.*, 1983, Geladi *et al.*, 1985), and commonly used in the initial stages of processing data for model development (Afseth and Kohler, 2012). MSC is a row-oriented transformative technique which normalises certain effects from spectral data such as the scaling and offset (baseline) effects. The application of MSC is a two-step process the first of which is an estimation of the correction coefficients and secondly the correction of the original spectrum, which are represented by the following equations (Figure 2.8).

[A] $$\boldsymbol{x}_{org} = b_0 + b_{ref,1} \cdot \boldsymbol{x}_{ref} + \boldsymbol{e}$$

[B] $$\boldsymbol{x}_{corr} = \frac{\boldsymbol{x}_{org} - b_0}{b_{ref,1}} = \boldsymbol{x}_{ref} + \frac{\boldsymbol{e}}{b_{ref,1}}$$

[C]

Constituent	Definition
$\boldsymbol{x}_{org}$	Original measured sample spectra
$\boldsymbol{x}_{ref}$	Reference spectra used in data preprocessing
$\boldsymbol{e}$	Un-modelled part of $\boldsymbol{x}_{org}$
$\boldsymbol{x}_{corr}$	Corrected spectra
$b_0 + b_{ref,1}$	Scalar parameters that vary for each sample

Figure 2.8: Multiplicative scatter correction equations – [A] Estimation of the correction coefficients [B] Recorded spectrum correction [C] Equation legend (Rinnan *et al.*, 2009)

The reference spectra for the equation can be derived from either the average spectrum of the calibration set or a generic reference spectrum. To determine the scalar parameters, sample spectrum have to first be plotted against a reference spectrum. The next step is to fit a line through the plotted points using least-squares regression fit through the plotted data, with the resulting intercept and slope of the line corresponding to the scalar parameters. Transformed MSC data has significantly reduced noise and is more linear when compared to the original data. The literature demonstrates that MSC can be utilised as spectral pre-processing technique across a range of platforms and processes, such as food/beverages, agriculture and biopharmaceuticals (Maleki *et al.*, 2007, Rinnan *et al.*, 2009, Clavaud *et al.*, 2013, Mou *et al.*, 2014).

2.3.2.4 Standard Normal Variate (SNV)

Standard normal variate (SNV), is a transformative normalisation technique similar to MSC (Figure 2.9) and has been demonstrated to be suitable when utilised in the pre-processing stage of some spectral data (Syvilay *et al.*, 2015).

[A] $$\boldsymbol{x}_{corr} = \frac{\boldsymbol{x}_{org} - a_0}{a_1}$$

[B]

Constituent	Definition
$\boldsymbol{x}_{corr}$	Corrected spectra
$\boldsymbol{x}_{org}$	Original measured sample spectra
a_0	Average value of the sample spectrum to be corrected
a_1	Standard deviation of the sample spectrum

Figure 2.9: Standard normal variate equations – [A] SNV correction [B] Equation legend (Rinnan *et al.*, 2009)

As stated SNV signal-correction works in a similar format to that utilised in MSC, however the main difference between these two methods is that a reference spectrum is not utilised in the dataset pre-processing. The alternative is that each sample undergoes the transformation individually, as opposed to applying the transformation to the whole data set at once (Guo *et al.*, 1999, Rinnan *et al.*, 2009, Varmuza and Filzmoser, 2016).

2.3.3 Partial Least Squares Modelling

Partial least squares (PLS) is a regression algorithm, used in process modelling which aims to correlate the spectral variance to measured analyte concentrations. It achieves this by integrating the identification of maximum spectral variation from principal component analysis (PCA), with the correlation of observed spectral features to known concentrations of analytes from multiple linear regression (MLR) (Rathore *et al.*, 2014). There are two methods of PLS that can be performed, PLS1 and PLS2. PLS1, constructs individual calibration models for each response (Y) variable. Alternatively PLS2, constructs only a single calibration model which incorporates all of the response variables simultaneously (Gemperline, 2006b).

When PLS is utilised it decomposes the data blocks into the corresponding scores and loadings matrices using the following formulas (Figure 2.10).

$$X = TP^T + E$$

$$Y = UQ^T + F$$

Figure 2.10: PLS decomposition equations

The X and Y matrices are respectively composed of, the *T* and *U* which are representative of the scores, while the loadings correspond to P^T and Q^T and *E*/*F* make up the residual matrices, which are approximated to be zero by the algorithm. However because the above equations calculate the latent vectors of the X and Y blocks independently, this can result in them only having a weak relation to one another. But this can be improved by calculating what is termed the 'inner relationship' between the scores of the X (T) and Y (U) blocks, by using a predictive formulation that also includes the matrix of the regression coefficients (W) (Figure 2.11) (Trygg and Wold, 1998, Abdi, 2003, Gemperline, 2006b).

$$U = TW$$

Figure 2.11: 'inner relationship' PLS equation

The advantage of utilising PLS over MLR for modelling is that it enables analysis of large numbers of predictor (X) variables which can be noisy or collinear, and can model multiple response (Y) variables at once (Wold *et al.*, 2001, Gemperline, 2006b, Rathore *et al.*, 2014, Varmuza and Filzmoser, 2016).

2.3.4 Evaluation & Validation of Models

2.3.4.1 Calibration

Overall quality of models is evaluated based upon the statistical values derived from the root mean square error of calibration (RMSEC), root mean square error of validation (RMSEV) and root mean square error of prediction (RMSEP) which should be as low as

possible. One indicator that a model has a good predictive ability is when the values for the RMSEC and RMSEP are similar. The value of the RMSEC is derived from a calculation of the standard deviation between the reference (actual) and spectroscopic (predicted) datasets, which are present in calibration dataset. This is calculated by using the equation in Figure 2.12, with n representing the number of samples and f the number of factors utilised in the model.

$$RMSEC = \sqrt{\frac{\sum_{i=1}^{n}(actual - predicted)^2}{n - f - 1}}$$

Figure 2.12: Root mean square error of calibration (RMSEC) calculation

Another indicator of good model predictive ability between the reference and predicted values is the coefficient of determination (R^2) values. R^2 values closest to one indicate that the two datasets are well correlated, with the data points well aligned to the straight line of the model (Roychoudhury *et al.*, 2006a, Kamruzzaman *et al.*, 2012, Sileoni *et al.*, 2013).

2.3.4.2 Internal Validation

The first stage of truly assessing model predictive performance is carrying out an internal validation, expressed as the RMSEV. One of the most commonly utilised internal validation techniques is leave one out cross validation. This technique operates by constructing models from the calibration data, where initially the first sample excluded and is then predicted. The strategy is then repeated for each sample separately until all of the calibration dataset samples have been tested (Brereton, 2007). The downside to this technique is that the data used to test the model was initially included in its construction. Therefore the model isn't tested as thoroughly as it would against data not with the initial

calibration dataset. So even with the internal validation statistics indicating good predictive ability the truest test of the model comes from an external validation.

2.3.4.3 External Validation

External validation is carried out by introducing data (validation dataset) not incorporated into the initial calibration dataset, which is where the RMSEP is used to indicate predictive ability of the model. This is expressed by the function in Figure 2.13, representing the standard deviation between the reference (actual) and validation (predicted) datasets and n which corresponds to the number of samples from the validation dataset (Roychoudhury *et al.*, 2006a, Sileoni *et al.*, 2013).

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n}(actual - predicted)^2}{n}}$$

Figure 2.13: Root mean square error of prediction (RMSEP) calculation

2.3.5 Design of Experiment

As previously outlined the purpose of chemometrics, is to reduce or remove data variation in the data prior to the construction of process models, through the use of various preprocessing techniques. However the selection of preprocessing techniques has historically been based on previous experience and knowledge. While this approach will most likely produce an acceptable model, it is most likely not to be the optimal approach. This is where a Design of Experiment (DoE) strategy can be applied to the selection of preprocessing techniques.

There are two basic principles by which DoE functions, firstly it systematically tests the effects that the predefined levels of the preprocessing techniques (variables) have upon the data (i.e. the RMSE value of either calibration or validation). Secondly, it also assesses

the interactions that each variable has on each other and the subsequent effects this has on the RMSE value. An example of spectral preprocessing using DoE can be seen in Figure 2.14 which utilises the DoEman software (CPACT, University of Strathclyde, Department of Pure & Applied Chemistry) (Flåten and Walmsley, 2003, Gerretzen *et al.*, 2015, Gerretzen *et al.*, 2016).

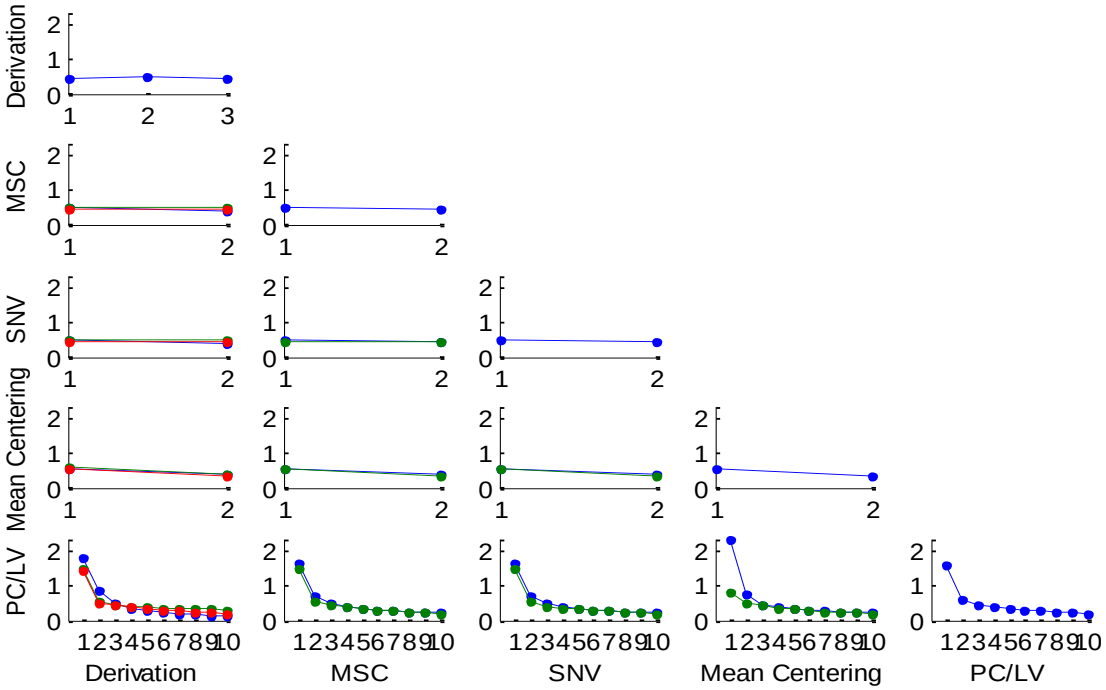


Figure 2.14: DoEman example plot. The main diagonal plots of the DoE matrix reflect main effects, off diagonal plots represent subsequent interaction between variables. Levels of the variables (columns) correspond to the different lines in the plots. Levels of the rows are found along the abscissas in the plots. Adapted from; (Flåten and Walmsley, 2003).

2.4 Data Fusion

Data fusion is a division of chemometrics, in which data from various analysis techniques is collected and analysed simultaneously. The aim of this technique is to extract more qualitative and quantitative information, than would be through individual analysis. In

terms of data modelling, application of data fusion has a number of advantages over individual analysis, such as reduction in errors and improvements in detection of previously obscured variables. These benefits can ultimately lead to better model predictions, when the data is subsequently modelled (Dearing *et al.*, 2011, Khaleghi *et al.*, 2013, Silvestri *et al.*, 2014). The 'simplest' form of data fusion is termed as low-level data fusion, which involves concatenation of the different datasets. The fused dataset is scaled, and then undergoes the normal preprocessing and modelling (Forshed *et al.*, 2007, Dearing *et al.*, 2011).

2.5 Chinese Hamster Ovary (CHO) Cells

Most of the current production of biopharmaceuticals is carried out using mammalian cell cultures. This is because they are capable of producing the proteins which possess the required posttranslational glycosylations, and are nearest in structure to authentic human proteins. These traits subsequently decrease the chance of consumers having an immunogenic response to the drug and deliver greater stability and efficacy of the drug (Del Val *et al.*, 2012).

One particular cell line that has been extensively used in the industrial production of biopharmaceuticals is the Chinese Hamster Ovary (CHO) cell line, initially isolated in the 1950's by Theodore Puck (Reichert, 2012, Bandaranayake and Almo, 2014). Production of the earliest clinically approved recombinant protein from mammalian cells, tissue plasminogen activator (tPA), was done using CHO cells (Kaufman *et al.*, 1985). Since then, use of CHO cells for production of therapeutic proteins is estimated to make up more than seventy per cent of the global market, with an annual sales value of one hundred billion U.S dollars and in the European Union, 47 out of 77 therapeutic EMA approved glycoprotein drugs are produced by CHO cells (Jayapal *et al.*, 2007, Del Val *et al.*, 2012, Bandaranayake and Almo, 2014).

Some significant developments in CHO cultures have arisen through media optimisation approaches, such as feeding strategies. This has in some cases led to increased antibody product yields ranging from 2-6 g/L. As a result the production of recombinant proteins at an industrial scale is largely carried out using fed batch systems. This is where concentrated feeds are introduced in small quantities over a period of time to improve cell growth, viability and production (Huang *et al.*, 2010, Bandaranayake and Almo, 2014). The adaptability of CHO cells is in part what allows for transition into chemically defined serum free media and growth at high densities, but this adaptability does have disadvantages. One disadvantage of note is that each desired product requires the selection of a clone or clones that exhibit specific properties, however this selection is not without fault as phenotypic drift can occur (Jadhav *et al.*, 2013).

CHO cells may consume certain nutrients such as glucose, at a rate which is greater than their stoichiometric needs and then produce lactate and ammonia as waste products (Dean and Reddy, 2013). The production of lactate has a detrimental effect on both growth and production of monoclonal antibodies (mAbs) in CHO cell lines and a number of methods have been tested to limit this, such as reduction of glucose concentration (Gagnon *et al.*, 2011). There are two distinct phases in CHO cell metabolism in a fed batch system the first of which is the growth phase. This phase is characterised by a relatively high rate of cellular growth and the production of lactate. The second phase is known as the stationary phase, in which the rate of cellular growth is reduced or has ceased and the production of recombinant proteins is increased (Dean and Reddy, 2013).

One key adaptability advantage of CHO cell lines is their genomic variability, which has allowed for manipulation and subsequent isolation of mutant lines with beneficial phenotypes. Often these mutants are designed with a selection marker such as nutrient dependency thus making them desirable for the development of producer lines (Del Val *et*

al., 2012). There are three engineering strategies utilised for CHO cell line development, genetic engineering, cellular engineering and metabolic engineering (Datta *et al.*, 2013).

The first strategy for CHO cell line development, genetic engineering, is where a specific gene is expressed for the production of the target product. This approach has a number of steps, firstly the target gene is cloned into an expression-vector. Next the vector is transfected and integrated into the CHO cells. Finally the most stable clones are selected and the expression level is optimised by gene amplification. (Jayapal *et al.*, 2007, Walsh, 2010, Datta *et al.*, 2013).

Cellular engineering is performed to alter phenotypes for improved growth and productivity (Datta *et al.*, 2013). The goal of this strategy is to optimise normal cellular processes, which can be achieved by engineering the cells to have a decreased production of lactate, resistance to apoptosis and improved glycosylation patterns (Jeon *et al.*, 2011). There are a number of methods by which the cellular processes of CHO can be engineered, one of which is utilising microRNAs (miRNAs) to down-regulate gene expression of groups of genes. These miRNA are small non-coding RNA molecules which carry out the down-regulating of genes by binding messenger RNA (mRNA) and thus inhibit translation of the gene protein (Datta *et al.*, 2013).

The final method for CHO cell engineering is metabolic engineering, which is used for the production of novel products. In order to achieve this, specific proteins in a metabolic pathway must be overexpressed and/or down-regulated (Wong *et al.*, 2006, Datta *et al.*, 2013). An example of this method was employed in the investigation into the feasibility of producing Heparin. This was done by exploiting metabolic pathways present in the CHO cells similar to that involved in Heparin production (Baik *et al.*, 2012).

The drive to improve production through CHO cell engineering resulted in a collaboration between University of Minnesota and the Bioprocessing Technology Institute

of Singapore in 2002, to sequence the genome of a number of CHO cell line. From this a subsequent sequencing study was established between the Consortium on Chinese Hamster Ovary Cell Genomics with the Society for Biological Engineers (SBE), whose aim was to identify specific genetic markers correlated to increased productivity (Bandaranayake and Almo, 2014).

2.6 Monoclonal Antibodies

2.6.1 Background

In 1975, Georges Köhler and César Milstein published a method for producing a novel hybridoma, a hybrid cell line combining a myeloma cell and an antibody-producing B cell. This new cell line was able to not only produce high volumes of specific antibodies, monoclonal (mAbs), but also be grown in culture (Galluzzi *et al.*, 2012, Liu, 2014, Skolnick, 2016).

Various sectors have had a massive impact with the advent of mAbs such as, clinical and medical research. There are two main reasons behind the impact of mAbs firstly, the high affinity selective binding ability, which allows two of the same antigen types to be bound simultaneously. The second is the increased stability that they display both, *in vivo* and *in vitro*. In comparison to polyclonal antibodies, derived from immunised animal serum, mAbs have increased specificity and reduced risk of contamination. As a result of this mAbs have also improved numerous diagnostic areas, such as immunohistochemistry and immunofluorescence (Galluzzi *et al.*, 2012, Ecker *et al.*, 2015, Skolnick, 2016, Treacy and Knight, 2016).

Despite the advantages that mAbs provide, with their increased use some disadvantages have been highlighted, particularly in relation to their complexity. The

extensive cloning required for production of Immunoglobulin G (IgG) mAbs with specific glycosylation patterning, have been shown to produce low titres initially and thus require more resource commitment to optimise the production process. Some of other issues associated with mAbs include: prolonged half-life, which is undesirable in some applications, and complex engineering for designing specific binding sites on the mAb (Razinkov *et al.*, 2015).

With the advent of monoclonal antibodies, from the Immunoglobulin G (IgG) family, they now represent one of the fastest growing biopharmaceutical products in terms of both, moving into the phase three clinical trials and subsequently gaining market approval (Liu, 2014, Reichert, 2014, Razinkov *et al.*, 2015). As a result of this development, the annual global sales for mAB based therapies is estimated at $75 billion, with fourty-seven mAbs currently approved in Europe the US for therapeutic use in humans for numerous medical issues and 300 mAb products in development (Liu, 2014, Ecker *et al.*, 2015).

2.6.2 mAb Structure

The structure of IgG mAbs (Figure 2.15) consists of four polypeptide chains of identical pairs of heavy and light chains linked with disulphide bonds. Each of the heavy chains is composed of three specific components: constant domains (C_H1, C_H2 and C_H3), the hinge region which is linked to the disulphide bonds and a variable domain (V_H). The light chains consist of two components; a constant (C_L) and variable (V_L) domain. IgG's overall structure can be placed into two distinct regions: firstly the fragment antigen binding (Fab) region, which comprises the V_L/C_L domains of the light chain and V_H/C_H1 domains from the heavy chain. The second is the fragment crystallisable (F_C) domain, which has the remaining C_H2 and C_H3 domains from the heavy chain. The variable domains within the Fab region are

what mediate the specificity of mAbs to antigens. Variable domains can be subdivided into the complementarity-determining region (CDR), responsible for binding of the antigen to the mAb, and a framework region for contact between the CDR to the antigen (Buss *et al.*, 2012, Ordás *et al.*, 2012, Treacy and Knight, 2016).

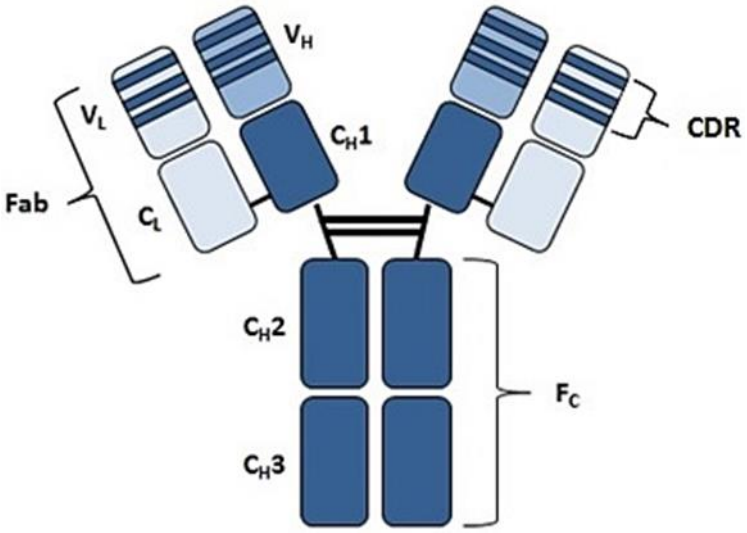


Figure 2.15: IgG Monoclonal Antibody Structure. Adapted from; (Buss *et al.*, 2012).

2.6.3 Categories of mAbs

There are four categories which that can be used to classify commercial mAbs, murine, chimaeric (murine/human), humanised and finally human (Figure 2.16). The first mAbs developed for clinical use were murine based, and as such resulted in a strong immunogenic response. In order to reduce the immunogenic response from murine mAbs, chimaeric and humanised antibodies were engineered to contain minimal animal based regions. Chimaeric antibodies are produced by, fusing the variable immunoglobulin (Ig) region of mouse antibody to the human constant Ig region. Humanised antibodies take the modification further, with the complementarity-determining regions (CDRs) of the chimeric mouse/human antibody were grafted into human antibody structure. The downside to humanised antibodies is that the engineering process is very time consuming, as it has to be

repeated for each desired antigen (Buss *et al.*, 2012, Glassy and Gupta, 2014, Treacy and Knight, 2016).

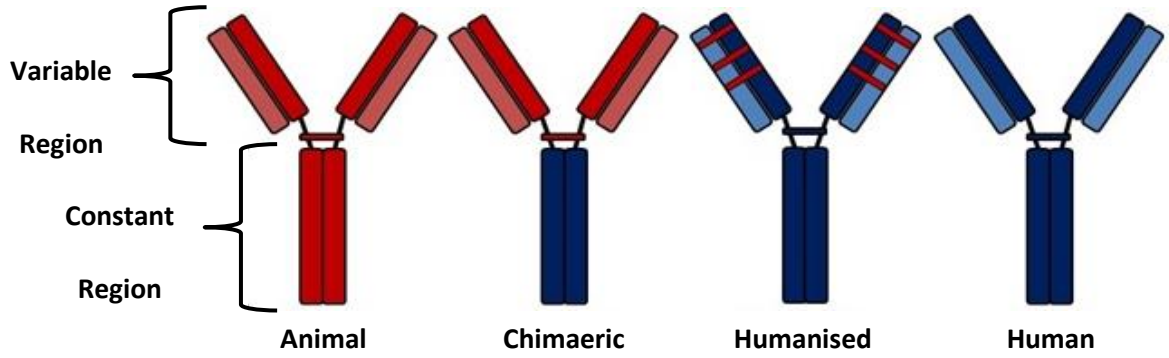


Figure 2.16: Categories of Monoclonal Antibody. Adapted from (Glassy and Gupta, 2014).

2.7 Aims

The main aim of this project is to look at approaches for progressing towards real-time monitoring and control of biomanufacturing processes, to this end application of the spectroscopic techniques NIR, MIR and Raman will be used at-line to construct process models. These techniques will be applied to CHO cell cultures for the purposes of monitoring key metabolites, cell density and product titre. The project will also focus on utilising low passage number cultures, to align with the needs of the industrial collaborators, and taking a DoE approach to spectral data preprocessing when forming the process models for each of the cell cultures. One other aspect that would be investigated would be the forming a single model from the combined data of all the techniques using data fusion. The models from each technique will be compared to the single "fused" model to determine which produces the lowest error of prediction. It is hypothesised that using a data fused model along with multiple analytical measurements, will result in a reduction in errors of prediction.

The potential impact and relevance of this project to the bioprocessing industry has three key aspects, firstly it aims to demonstrate the feasibility of triple spectrometric 'toolset' in moving towards on-line real time monitoring, as a viable alternative to more traditional at- and off-line monitoring techniques. The potential from utilising these techniques would be the ability to monitor a process as it is happening in real-time and be able to intervene at the first instance of an irregularity thus ensuring the integrity of the end product. Secondly this research seeks to directly aid the biopharmaceutical industry by carrying out research in partnership with two leading companies. Finally the utilisation of data fusion in complex biological processes, on which to date there has been limited published data to the author's knowledge, could in theory provide a more accurate means for developing real-time monitoring of biomanufacturing processes. Overall these aims are offering novel solutions to meet the demanding standards of the ever increasing bioprocess industry.

Chapter 3:

Materials & Methods

3.1 CHO Cell Lines

There were three Chinese Hamster Ovary cell lines provided for use in this study, which have been recorded as producing varying titres of a human Immunoglobulin G (hIgG) monoclonal antibody (mAb), cB72.3 (Table 3.1). These cell lines were kindly provided by Lonza (Slough, UK).

Table 3.1: Lonza CHO cell line predicted mAb production. Colour indicates the different concentrations of production, with the light to dark representing low to high.

Cell Line	**Product Concentration (g/L)**		
	Batch Shake Flask	Fed-batch Shake Flask	Fed-batch 10 L Bioreactor
124	0.3	1	1.5
47	0.8	2	2
42	0.5	2.5	3

As the experimental design for this work was focused on lab scale bioreactor culturing with industrial relevance, CHO cell line 42 was chosen as its predicted product titre was the highest.

3.2 Media & Feed Supplements

All media and feed supplements used in this work are chemically defined and animal origin-free, containing no hydrolysates, proteins or components which do not have a fully defined composition. These were selected for a number of reasons, industrial relevance, indicated use within the literature and also the interests expressed by the company supplying them, Thermo Fisher Scientific. This last point in particular was of great importance, as the industrial was interested in knowing the effects of low passage number cultures in these media. However all the specific composition of the media and supplements used were

confidential and not disclosed to the author. While overall composition of the three media was not known to the author some analysis indicated that initial glucose concentration was different across the three. The CD CHO medium contained the most starting glucose, while CD OptiCHO and Dynamis were more similar to each other ,but contained less than CD CHO.

3.2.1 Media

3.2.1.1 CD CHO

GIBCO CD CHO Medium (Life Technologies Ltd., Paisley, UK) is designed for suspension culture of Chinese Hamster Ovary (CHO) cells and the production of recombinant proteins, in batch and fed-batch cultures.

3.2.1.2 CD OptiCHO

GIBCO CD OptiCHO Medium (Life Technologies Ltd., Paisley, UK) has been further developed from CD CHO for growth of CHO cells and has been shown to increase product titre in fed-batch cultures.

3.2.1.3 Dynamis

GIBCO Dynamis Medium (Life Technologies Ltd., Paisley, UK) has been designed to produce the highest cell densities and product titres in both batch and fed-batch cultures of the three GIBCO media. Dynamis has had the glucose concentration formulated in order to minimise negative effects of lactic acid accumulation.

3.2.2 Feed Supplements

These feed supplements are utilised to extend standard batch cultures, by feeding them with concentrated glucose and essential nutrients over time.

3.2.2.1 CHO CD EfficientFeed™ A Liquid Nutrient Supplement

GIBCO CHO CD EfficientFeed™ A (EFA) liquid nutrient supplement (Life Technologies Ltd., Paisley, UK) is designed for a CD OptiCHO media fed-batch process.

The feeding strategy utilised for the CD OptiCHO cell cultures was taken and scaled up from one described by (Reinhart *et al.*, 2015a). They outline a feeding strategy utilising an IgG producing CHO cell line culture in flasks. The strategy required the addition of 10% v/v of EFA on days 3, 5, 7, 9 (40% total) to maintain the glucose concentration above 3 g/L.

3.2.2.2 CHO CD EfficientFeed™ B Liquid Nutrient Supplement

GIBCO CHO CD EfficientFeed™ B (EFB) liquid nutrient supplement (Life Technologies Ltd., Paisley, UK) is designed for fed-batch culture systems and is compatible with CD CHO culture media.

The CD CHO cell cultures were fed utilising a strategy based on one described by (Barrett *et al.*, 2012). They outline a feeding strategy utilising an IgG producing CHO cell line culture in a 5L bioreactor. To maintain the glucose concentration above 2 g/L, the culture was fed with 5.6% v/v EFB on days 3, 5, 7, 9, 11, 13, and 15 (39.2% total).

3.3 CHO Cell Line Development

The initial cell line vials provided by Lonza were in a CD CHO based freezing medium and labelled as passage number one, which formed the base for all further passage number development. The revival and seeding procedures that follow account for scaling through an initial passage number of four at the revival, through to the bioreactor at a final passage number of seven, for all processes.

3.3.1 Cell Banking

CHO cell cultures were spun down at 1000 rpm for 5 min and the supernatant was removed (Thermo Fisher Scientific, UK, Model: Heraeus Megafuge 8R Centrifuge). To cryopreserve the resulting cell pellet at -80 ^{o}C, 'freezing' medium composed of 90 % v/v CD CHO medium and 10 % v/v DMSO was added to resuspend the cells, to a final density of 1x10^{7} cells/mL. The cells were evenly suspended by gently aspirating with a pipette, once this was carried out 1.8 mL of the suspension was transferred into cryovials. Vials were placed into a Nalgene 'Mr Frosty' (Sigma-Aldrich Company Ltd., Dorset, UK) container containing isopropanol, allowing gradual freezing of cells and minimising cellular damage, and frozen for a week at -80 ^{o}C. Vials designated for the master cell bank were transferred to liquid nitrogen storage for long term preservation, while working cell bank cultures were maintained at -80 ^{o}C.

3.3.2 Cell Revival

A CHO cell culture vial was thawed at room temperature, to reduce thermal shock, while a bottle of appropriate media supplemented with 1 mL /L 25 mM L-methionine sulfoximine (MSX) (Sigma-Aldrich Company Ltd., Dorset, UK) was warmed to 37 ^{o}C. Once the culture was completely defrosted it was transferred to a falcon tube containing 10 mL of medium and spun down at 1000 rpm for 5 min (Thermo Fisher Scientific, UK, Model: Heraeus Megafuge 8R Centrifuge). The supernatant was removed and 10 mL of fresh medium was added to the cell pellet, then gently aspirated with a pipette to obtain a homogenous cell suspension. The cell suspension was then transferred to a 125 mL Erlenmeyer flask, with a vented cap and baffled base, and brought up to 30 mL using the media. Flasks were then incubated (Infors UK, Reigate, UK, Model: Minitron) at; 5% CO_2, 37 ^{o}C, 150 rpm and after 3

days of incubation flask contents were spun down at the previously stated parameters, the cell pellet was then resuspended in 30 mL of fresh media.

3.3.3 Seeding Procedure

The seed was developed from the initial revived culture in the appropriate media with the addition of MSX at the same conditions previously stated (Section 3.3.2), with cell density and viability calculated each day (Section 3.8.3). Once the culture reached a density of $3x10^6$ cells/mL, it was passaged and scaled up into a 250 mL Erlenmeyer shake flask, with a working volume of 150 mL and cell density of $0.3x10^6$ cells/mL. This culture was again grown until it reached a density of $3x10^6$ cells/mL and was scaled up into a 1 L Erlenmeyer shake flask, with a working volume of 300 mL until the cell density was $0.3x10^6$ cells/mL. The seed was ready to be utilised in the bioreactor vessel when it reached a density of $\geq 3x10^6$ cells/mL.

3.3.4 Production Procedure

For both batch and fed-batch cultures, an initial seed density of $0.3x\ 10^6$ cells/mL was used as the seed culture for the production vessel which was set up under the following conditions; pH 7, 40 % dO_2, 37 ^oC, 200 rpm. The aeration of the production vessels was maintained at 0.05 vvm and they were also supplemented with 0.1 ml/L of 10% Antifoam C (Sigma-Aldrich Company Ltd., Dorset, UK) when excessive foaming was observed. A 20 mL sample was taken every 24 hours for analysis, until the percentage viability was zero, at which point the run was ended. Feeding protocols for the fed-batch cultures, were unique to each process based on the basal media used (3.2.2).

3.4 Bioreactor

3.4.1 Applikon Bioreactor System

All experimental work was carried out using an Applikon bioreactor system consisting of a 7 L glass bioreactor, ADI 1010/1025 controller units and gas analyser (Figure 3.1).

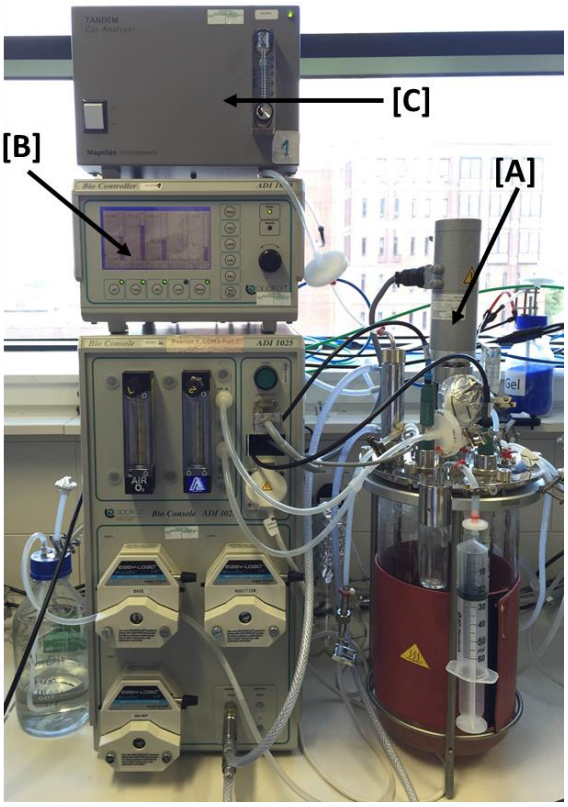


Figure 3.1: Applikon Bioreactor System. [A] Bioreactor Vessel [B] ADI 1010/1025 [C] Gas Analyser

3.4.1.1 Vessel

The vessel used was an Applikon 7 L dished bottom autoclavable glass bioreactor with a working volume of 5 L. This bioreactor has a head plate consisting of numerous ports for sampling and addition of titrants, as well as the various probes (pH, temperature and dissolved oxygen). The stirring was maintained by a single Rushton turbine impeller, while

aeration was provided by a ring sparger and to control the temperature a silicone heating blanket was wrapped around the vessel (Figure 3.2) [Applikon Biotechnology, Tewkesbury, UK].

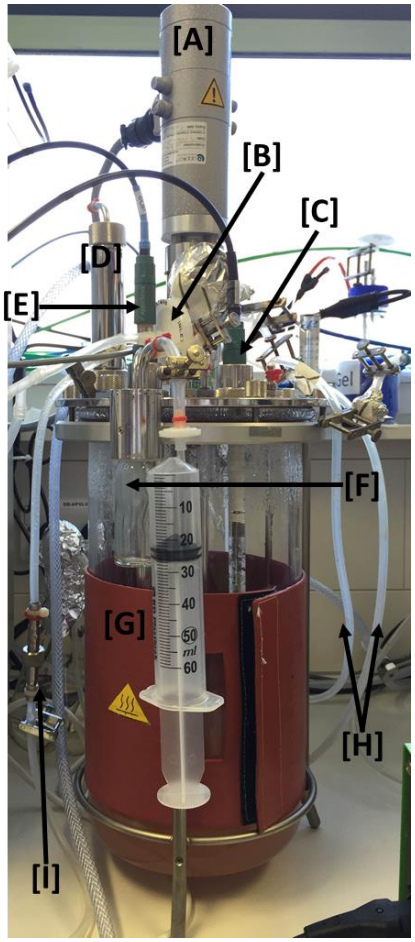


Figure 3.2: Applikon Bioreactor Vessel. [A] Motor [B] Gas Inlet [C] pH Probe [D] Condenser [E] Dissolve Oxygen Probe [F] Sampling Vial [G] Heating Jacket [H] Media & Feed Line [I] Base Line

3.4.1.2 Applikon BioController ADI 1010 & Applikon BioConsole ADI 1025

The controller is used to set, monitor and maintain the process variables of, pH, temperature, dissolved oxygen (dO_2) and stirrer speed, throughout the course of the run [Applikon Biotechnology, Tewkesbury, UK, Model: ADI 1010]. The ADI 1025 is an actuator

console coupled with the BioController ADI 1010, which enables the regulation of gas flow, temperature control, the titrant pumps and finally the flow of cold water to the condenser (Figure 3.3).

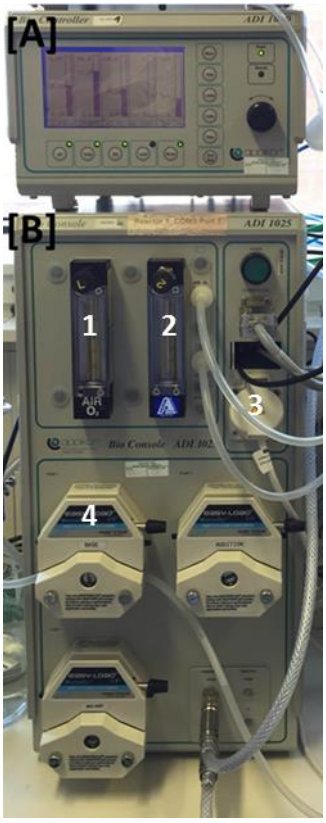


Figure 3.3: Applikon Control Units. [A] BioController ADI 1010 [B] Actuator Console ADI 1025: (1) Air/O_2 Rotameter (2) CO_2 Rotameter (3) Heating Jacket Power Supply (4) Titrant Pump

3.4.1.3 Off Gas Analyser

Analysis of the respiratory gases was carried out using a, Tandem Off Gas Analyser. This monitors the oxygen (O_2) uptake rate and carbon dioxide (CO_2) evolution rate of the gases being fed to and removed from the fermenter, using polyurethane tubing (SHPI, Taiwan) (Figure 3.1 [C]).

3.4.2 BioXpert

The BioXpert program is used in conjunction with Applikon bioreactors as a means of controlling processing set points and collecting process variable data. The program presents the data collected in either graphically or in a table and is exportable for further analysis [Applikon Biotechnology, Tewkesbury, UK].

3.5 NIR

3.5.1 Foss 6500

Near infrared spectral data was acquired using an at-line 6500 vis-NIR spectrophotometer (Foss NIRSystems, Maryland, USA) (Figure 3.4), with a spectral range of 1100 to 2500 nm and the spectral data gathered was displayed using Vision Near-Infrared Spectral Analysis Software (Foss NIRSystems, Maryland, USA, Version: 3.4).

Figure 3.4: At-line 6500 vis-NIR spectrophotometer. Highlighted region indicates the internal cuvette holder where the air reference and cuvettes loaded with the sample are scanned.

The samples used (1 mL) were all liquid based and were analysed in a 0.5 mm pathlength cuvette in transmission mode, with air as a reference. Transmission mode was selected as the samples of medium and cell culture scanned were optically clear and not overly viscous. The cuvette was cleaned before and after each sample using 70% ethanol to remove any matter within the viewing pane, washed with distilled water to remove any ethanol and then dried to ensure no trace of water/ethanol remained to interfere with the results of the scan. Samples were pipetted, undiluted, into the cuvette until the viewing pane was filled and placed into the holder within the spectrophotometer. Thirty-two scans of the sample were carried out, as this was a common number of accumulations used in the literature, and averaged to produce the final spectra which was referenced against the air background. This process was performed in triplicate to provide the most accurate representation of the sample and the results were consolidated and exported to Matlab for analysis.

3.6 MIR

3.6.1 MB3000 FT-IR with ATR Diamond Probe

The instrument that will be used to gather the MIR spectral data will be the at-line MB3000 FT-IR (ABB Inc. Analytical Measurements, Quebec, Canada) with a ATR Diamond Probe (Art Photonics, Berlin, Germany) covering a spectral range of 900 to 1200 cm^{-1} (Figure 3.5).

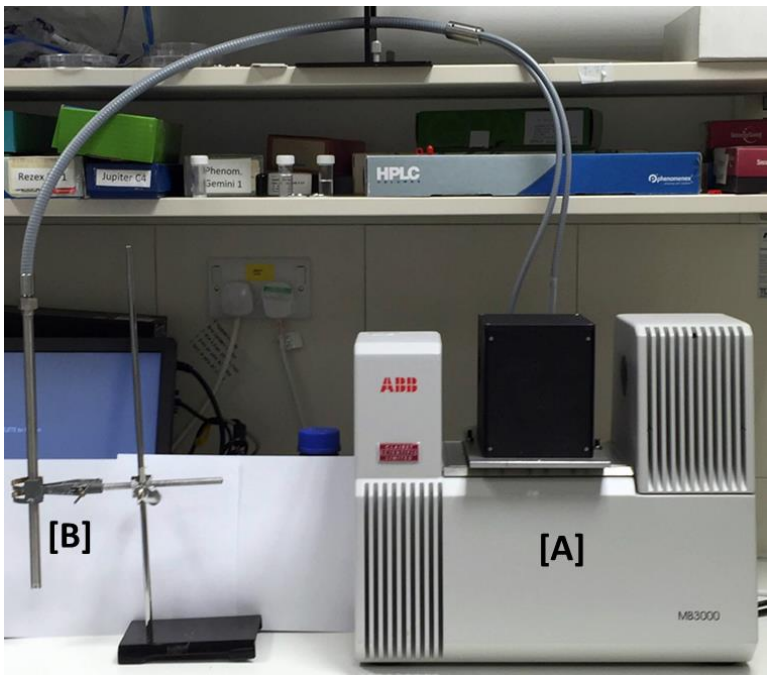


Figure 3.5: At-line ABB MB3000 FT-IR system with ATR Diamond Probe. [A] MB3000 FT-IR unit with modified probe attachment [B] ATR Diamond Probe.

To analyse the sample, a 3 mL undiluted volume was held in a 5 ml vial and the probe was lowered into the vial ensuring that there are no air bubbles present until the diamond was immersed in the sample. The probe head was cleaned before and after each sample using 70% ethanol and then washed with distilled water before being dried to ensure there was no trace of water/ethanol that might interfere with the results. As with the NIR (3.5) analysis all of the samples were scanned thirty-two times, then averaged and carried out in triplicate to ensure the most accurate result was obtained before being exported to Matlab for analysis.

3.7 Raman

3.7.1 RamanRXN2 Analyzer

The at-line RamanRXN2 Analyzer (Kaiser Optical Systems Inc., Michigan, USA) was used for Raman spectrometric analysis. This instrument utilises the Mk II Probe Head (Kaiser Optical Systems Inc., Michigan, USA) fitted onto an enclosed sample compartment (Kaiser Optical Systems Inc., Michigan, USA, Model: RXN-ESC) allowing at-line scanning of the samples at a wavelength of 785 nm over a spectral range of 100 to 3500 cm^{-1} (Figure 3.6).

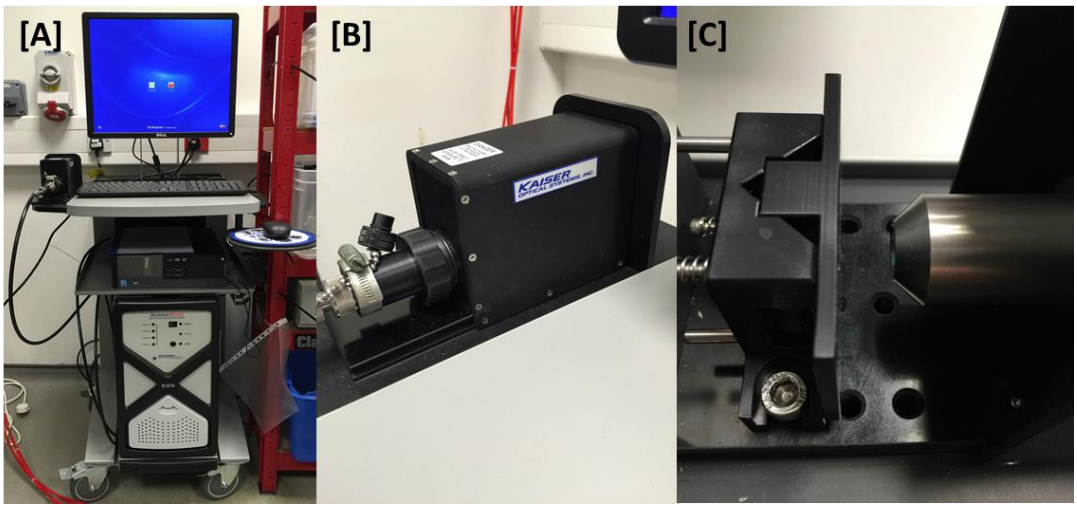


Figure 3.6: At-line RamanRXN2 Analyzer. Highlighted regions indicate [A] At-line RamanRXN2 Analyzer [B] at-line enclosed sample compartment [C] Mk II Probe Head

The iC Raman software package (Kaiser Optical Systems Inc., Michigan, USA) was used to control the RamanRXN2 spectrometer and provide basic data presentation, it will also be used to communicate with other data analysis packages, such as Matlab. Samples were analysed by placing a Kimble™ scintillation vial containing a 5 mL undiluted sample into the enclosed sample compartment with the Mk II Probe Head attached and initiating the scanning. To ensure no interference from previous samples the probe and vial was cleaned with 70% ethanol, followed by distilled water and then dried, before and after each

sample was scanned. Samples were analysed in the same manner as was used for the previous spectroscopic techniques (3.5 & 3.6).

3.8 Spectral Analysis

3.8.1 Reference Spectra

As is discussed in the results chapters the spectral regions focused upon were based on observations from both the published literature and upon previously performed experimental analysis on pure media as blanks and component samples. The following reference spectra represent the observations previously noted by the author.

3.8.1.1 NIR Reference Spectra

The initial NIR observations (Figure 3.7) indicated that in order to determine spectral regions of focus some basic processing was required to visualise the activity and reduce the effect of water upon the spectra. By applying a Savitzky-golay derivatisation to the data, clear activity in the regions of interest indicated by the literature(2200 to 2400 nm) could be seen (Figure 3.8) (Chen *et al.*, 2004, Goodarzi *et al.*, 2015).

The observations demonstrate that while the three media used in this research have some similarities, in terms of specific peaks, there are also some subtle differences which are probably a result of the specific media compositions. In comparison the pure component spectra of glucose shares a similar overall structure to the three media, indicating the presence and influence of glucose on the spectra of the media. In contrast the distinct lactate peaks are absent in the pure media samples. When comparing the synthetic glucose/lactate mix NIR spectra to process samples (Chapter 4, 5 & 6) these appear to be similar in structure. As this specific spectral structure is only present when

both glucose and lactate are present, it can be used as a means to determine the concentrations of both metabolites in the process samples.

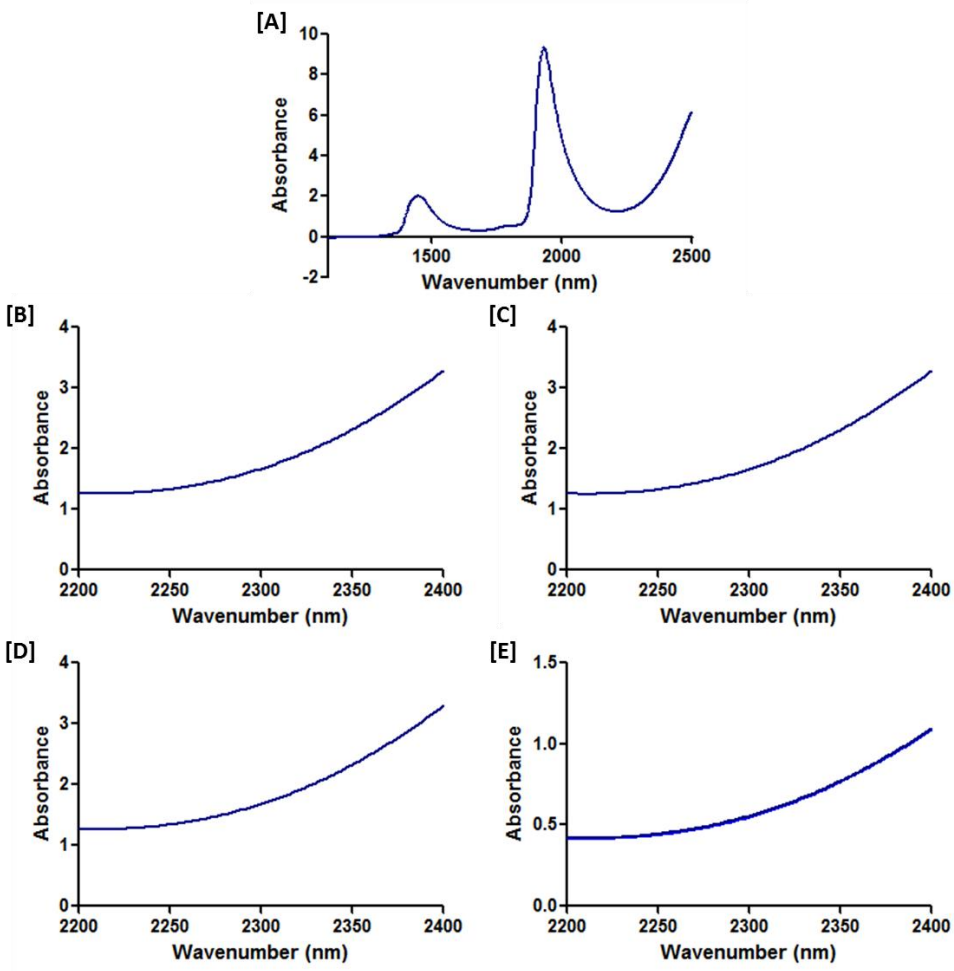


Figure 3.7: Raw Reference NIR Spectral Data – Whole Spectra & Target Region. [A] Standard Whole Spectra [B] CD CHO medium [C] CD OptiCHO medium [D] Dynamis medium [E] Pure Component spectra of Lactate(Light Blue), Glucose (Dark Blue) & Glucose/Lactate mix (Darkest Blue)

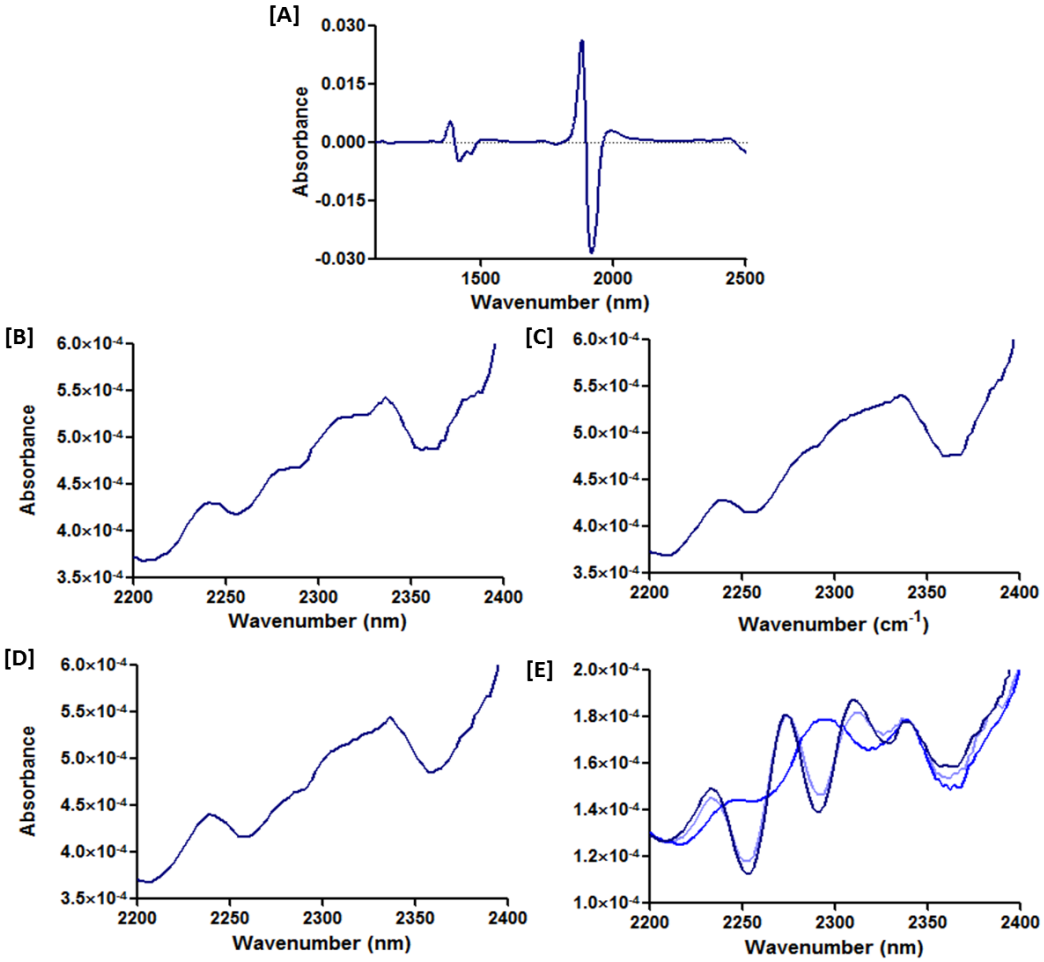


Figure 3.8: Processed Reference NIR Spectral Data – Whole Spectra & Target Region. [A] Standard Whole Spectra [B] CD CHO medium [C] CD OptiCHO medium [D] Dynamis medium [E] Pure Component spectra of Lactate (Light Blue), Glucose (Dark Blue) & Glucose/Lactate mix (Darkest Blue).

3.8.1.2 MIR Reference Spectra

Previously performed MIR analysis on pure component glucose and lactate, at a wavenumber range of 900 to 1200 cm^{-1} was used to initially identify the spectral signals of interest (Figure 3.9), the results of which correlated to those region identified by the literature (Rhiel *et al.*, 2002, Lu *et al.*, 2011). This analysis was carried out on a Zinc Selenide (ZnSe) trough plate Thermo Nicolet AVATAR 370 FTIR system. As this system was not

available at the time of the present study, as it needed repairs, the system outlined in section 3.6.1 was used and the previous data generated from the AVATAR 370 FTIR system was utilised as a guide for further analysis. In contrast to the previous system, the current system used in this study didn't display any particular signal at the previously identified regions (Figure 3.10). However by applying the DoE strategy to the spectral preprocessing, it was hoped that the previously identified signals would be elucidated in the data from the current MIR system.

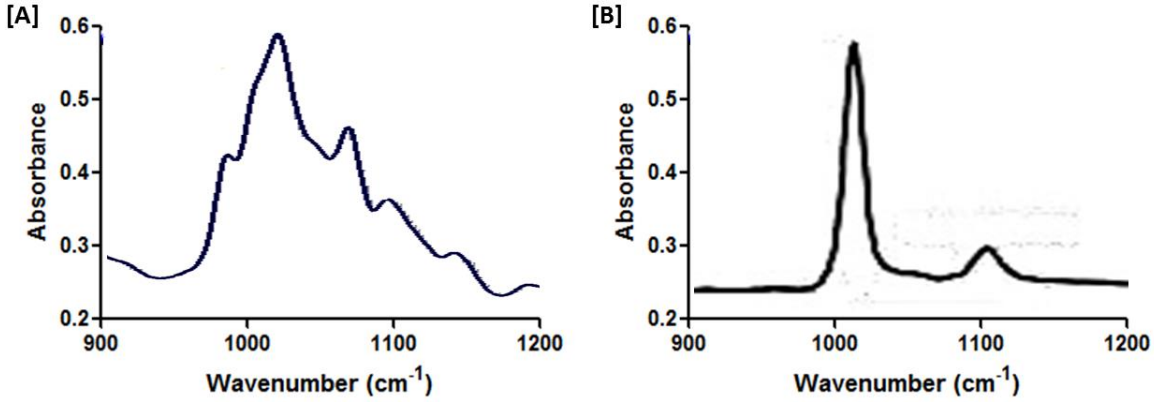


Figure 3.9: Thermo Nicolet AVATAR 370 FTIR Raw Reference Spectral Data of Target Region.

[A] Pure Component spectra of Glucose [B] Pure Component spectra of Lactate

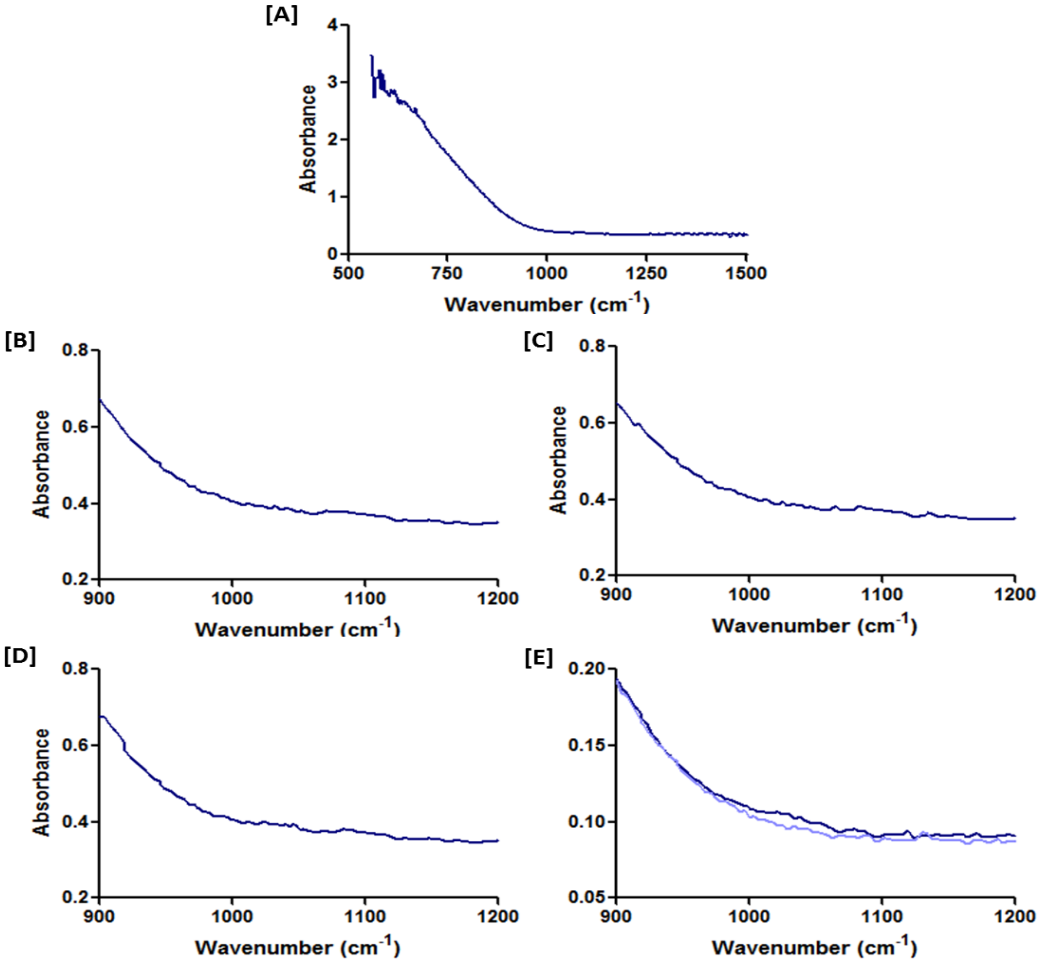


Figure 3.10: At-line ABB MB3000 FT-IR Raw Reference MIR Spectral Data – Whole Spectra & Target Region. [A] Standard Whole Spectra [B] CD CHO medium [C] CD OptiCHO medium [D] Dynamis medium [E] Pure Component spectra of Lactate (Light Blue) & Glucose (Dark Blue)

3.8.1.3 Raman Reference Spectra

The Raman analysis (Figure 3.11) indicated that all three media had some spectral activity in the region the literature indicated glucose was present (1100 to 1150 cm^{-1}) (Mehdizadeh *et al.*, 2015, Singh *et al.*, 2015). That these peaks were of varying intensities also confirmed previous knowledge of the variance of starting glucose concentration between the media. The pure component glucose spectral also showed activity in the same region as the media,

while pure component lactate activity was observed in the region highlighted by the literature (830 to 860 cm^{-1}).

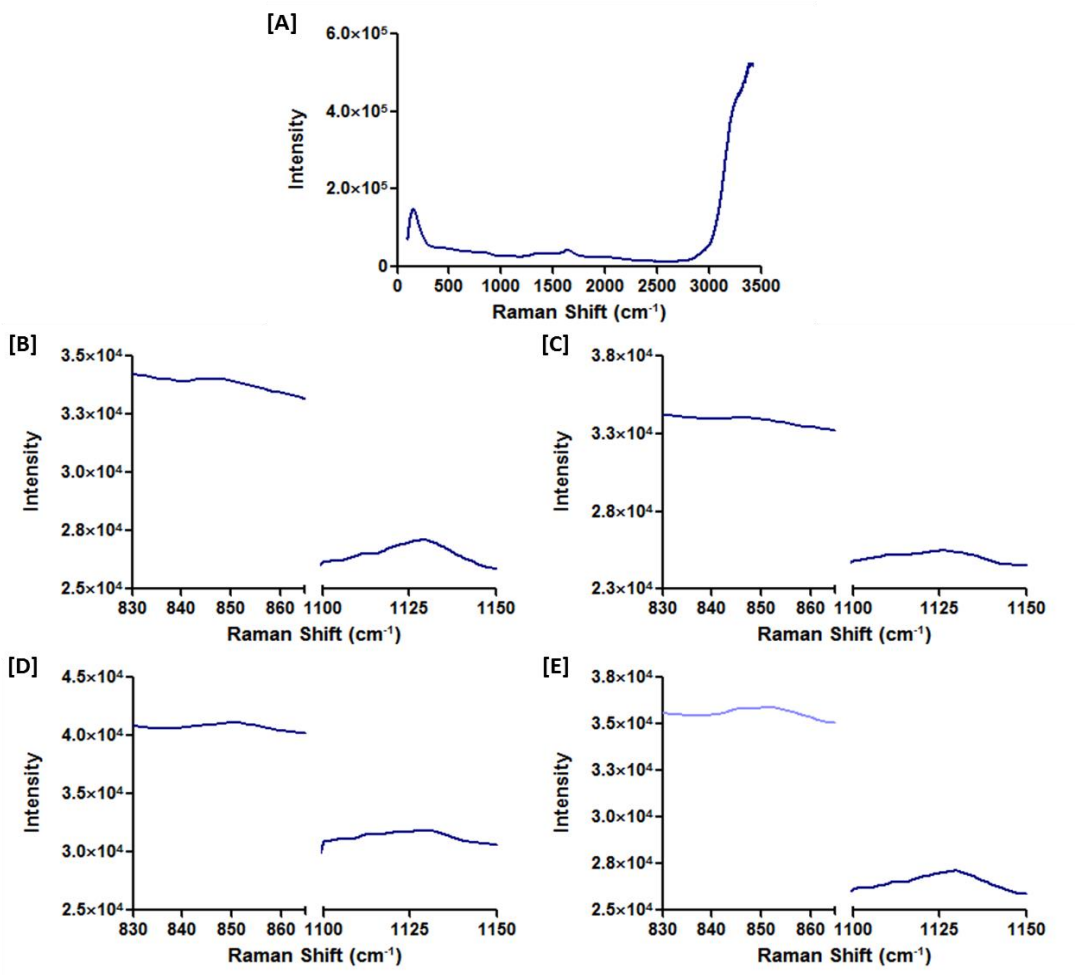


Figure 3.11: Raw Raman Reference Spectral Data – Whole Spectra & Target Regions. [A] Standard Whole Spectra [B] CD CHO medium [C] CD OptiCHO medium [D] Dynamis medium [E] Pure Component spectra of Lactate (Light Blue) & Glucose (Dark Blue)

To calibrate the instrument and determine the optimal focal point for sample analysis, cyclohexane was used as a reference due to its distinct and easily identifiable Raman spectral signature (Figure 3.12).

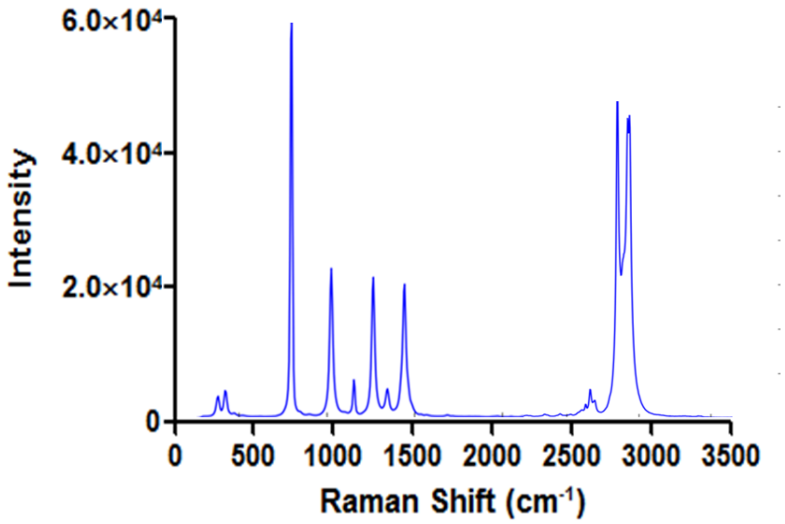


Figure 3.12: Raw Raman Cyclohexane Reference Spectral Data

3.8.2 Matlab

All spectral data was analysed using the Matlab analysis software version R2012b (Natick, MA, USA), which included the additional statistical processing 'toolboxes' for PCA, PLS and DoE analysis of the spectral data.

3.8.3 DoEman

As described in 2.3.5 DoEman software (CPACT, University of Strathclyde, Department of Pure & Applied Chemistry) was utilised in determining the optimal preprocessing strategy for all the spectral data sets. The spectral and reference data is firstly loaded into Matlab before it is uploaded in the DoEman software. Once loaded into DoEman, there are a number of options that can be selected for investigation; regression method, no. of components and preprocessing techniques. After DoEman has completed the analysis it will generate a matrix of plots, as seen in the aforementioned section (Figure 2.14), from this the optimal preprocessing strategy can be selected based on the lowest RMSE. For this

research the pre-processing techniques investigated in Chapters 4, 5 and 6 can be seen in Table 3.2.

Table 3.2: DoE parameters and levels tested in the pre-processing of the spectral data in all media processes, for glucose and lactate.

Parameter	Levels
Regression	PLS 1
	PLS 2
Savitzky-Golay	None
	1st Derivative
	2nd Derivative
Scatter Correction	None
	MSC
Normalisation	None
	SNV
Scaling	None
	Mean Centering
No. of Latent Variables	1 - 10

3.9 Reference Analysis

3.9.1 YSI: Glucose, Lactate, Glutamine, Glutamate & Ammonia Analyses

To gather reference data from the various metabolites a YSI 2950 Biochemistry Analyzer (YSI Ltd. Hampshire, UK) was used (Figure 3.13). All the metabolite analysis was done in triplicate by programing the instrument to take three 100 µL volumes, from 1 mL process samples that were loaded, and then mixed with the appropriate buffer/calibrator (Figure

3.13 [B]). These samples were then introduced to the enzyme sensor probes, for metabolite measurement.

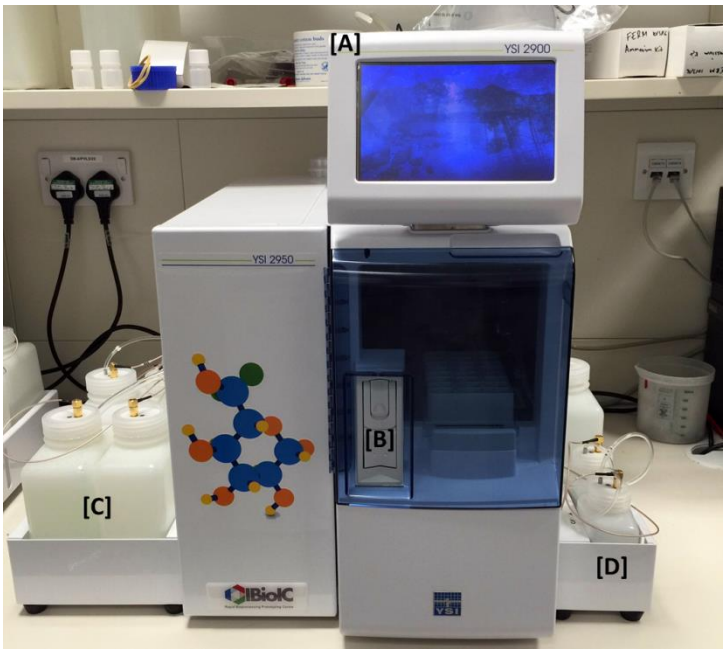


Figure 3.13: YSI 2950 Biochemistry Analyzer. [A] Program & Result Display [B] Sample Holder & Analysis Chamber [C] Buffer & Waste Bottles [D] Calibrator Standard Bottles

This instrument utilises enzyme sensor probes to quantify the specific metabolites within a sample. Each probe is fitted with a three-layer membrane which covers the face electrode and Figure 3.14 shows a schematic of each membrane layer.

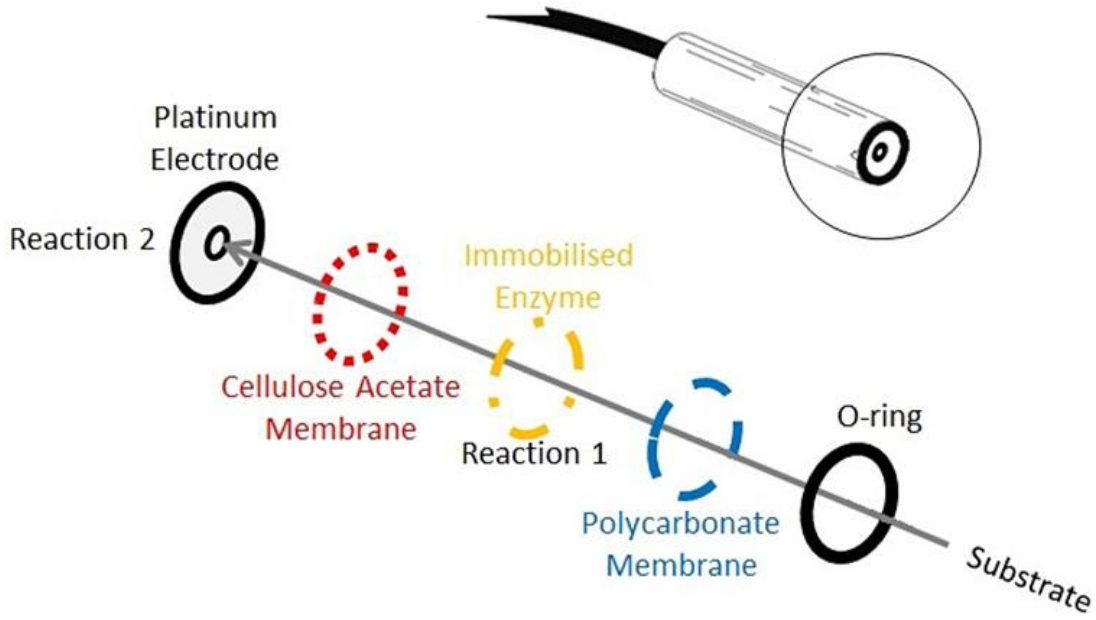


Figure 3.14: YSI Sensor Probe & Membrane

The face of the probe, covered by the membrane, facilitates diffusion of the substrate and when it contacts the immobilized enzyme it is rapidly oxidized, producing hydrogen peroxide (Figure 3.14 - Reaction 1). The hydrogen peroxide (H_2O_2) is, in turn, oxidized at the platinum anode, producing electrons (Figure 3.14 – Reaction 2). A dynamic equilibrium is achieved when the rate of H_2O_2 production and the rate at which H_2O_2 leaves the immobilized enzyme layer are constant. The electron flow is linearly proportional to the steady state H_2O_2 concentration and therefore, to the concentration of the substrate. The membranes, buffers and calibrators required to analyse each metabolite are outlined in Table 3.3.

Table 3.3: Metabolite YSI membranes, buffers and calibrators

Metabolite	Membrane	Buffer	Calibrators
Glucose	YSI 2365	YSI 2357	YSI 2776
(L-)Lactate	YSI 2329		
Glutamine	YSI 2735		YSI 2736
Glutamate	YSI 2754		YSI 2755
Ammonia	YSI 2974 (Probe)	YSI 2970	YSI 2972

3.9.2 ELISA: Product Titre

Quantification of human IgG (hIgG) produced in the cell culture processes were carried out using a hIgG 'sandwich' ELISA Kit (Bethyl Laboratories, Inc., Montgomery, TX, USA) (Figure 3.15).

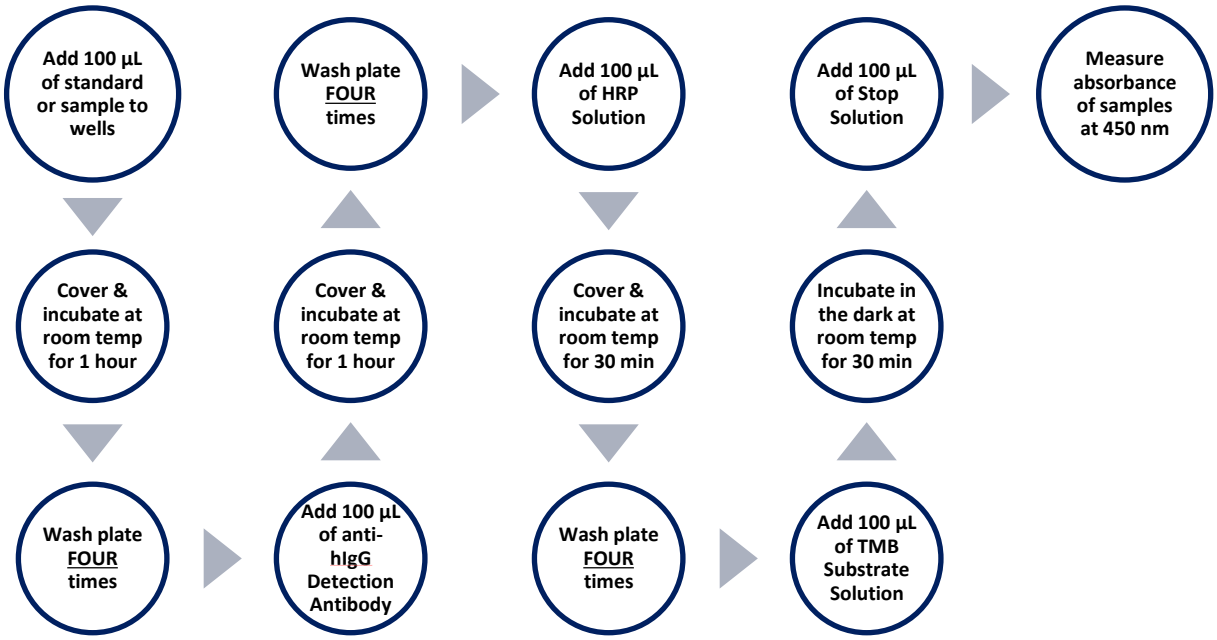


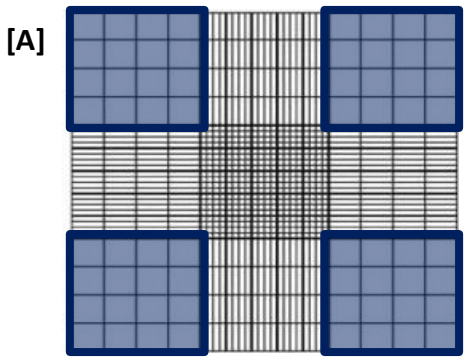
Figure 3.15: Step by step protocol for hIgG ELISA quantification

The hIgG in the 100 µL samples are bound to the wells by anti-hIgG antibody and any unbound proteins were washed off. A 100 µL of biotinylated detection antibody was added, to bind to the hIgG bound to the well, creating the 'sandwich'. Horseradish

peroxidase (100 µL) was then added to catalyse a colorimetric reaction, producing a blue colour which then becomes yellow upon termination by the addition of 100 µL sulfuric acid. Absorbance of the well contents was read at 450 nm, with the intensity being directly related to the concentration of hIgG present in the samples and hIgG concentrations are determined by generating a standard curve from the standards provided in the kit.

3.9.3 Cell Density & Viable Cell Counting

This was carried out using a haemocytometer (Improved Neubauer, Hawksley, West Sussex, UK). The haemocytometer and cover-slip were first cleaned with 70% ethanol and the cover-slip was then affixed to the haemocytometer. 100 µL of the culture, well mixed by gently aspirating was added to 100 µL of 0.4% Trypan Blue stain. Approximately 10µl of the trypan blue/cell mix was pipetted onto the haemocytometer and as Trypan Blue is a "vital stain" it excludes live (viable) cells, therefore only dead cells become stained (Insert figure of slide with dead/live cells). To calculate the cell density, four corner sections from the haemocytometer grid were used and incorporated into the equation, as highlighted in Figure 3.16.



$$No.Cells/mL = \left(\frac{\frac{No.Cells\ Counted}{No.Squares\ Counted} \times Dilution\ Factor}{Volume\ of\ Square} \right) \times 1000$$

Figure 3.16: Cell counting procedure – [A] Improved Neubauer Haemocytometer Layout with cells counted highlighted in blue [B] Cell density calculation

Chapter 4:

A Multisensor At-Line Approach to Monitor Chinese Hamster Ovary Cell Line in a CD CHO Batch and Fed Batch Culture System

4.1 Overview

The bioprocesses that were investigated in this research were batch and then fed batch cultures of the industrial CHO 42 mAb (hIgG) producing cell line, at a low passage number, in the commercial CD CHO medium. These were chosen as the literature indicated these processes would provide both a suitable baseline and the first developmental stage for the overall research program (Barrett *et al.*, 2012, Ling *et al.*, 2015). Due to the ever increasing regulatory stance and high standards set out by the PAT, QbD and ICHQ8 (R2) initiatives for biopharmaceutical production, progressing towards real time monitoring of these processes has never been more vital. It was for these reasons that application of the spectroscopic techniques NIR, MIR and Raman was chosen for process monitoring.

The replicates for both of the processes were run under the parameters set out in Section 3.3, as was the sampling procedure carried out over the course of the processes. To monitor the progress of the batches *in-situ* measurements of pH, dissolved oxygen and temperature were made, while cell density and viability were analysed at-line immediately when samples were taken. The metabolite and product analysis was split into two categories, reference (offline) or modelling (at-line). All reference analysis was obtained from the YSI and ELISA, while modelling data was gathered from the NIR, MIR and Raman spectrometers.

Spectroscopic data collected contained information relating to not only the target metabolites but also numerous other components within the sample matrix. This influence from the other components, which is reflected in the spectra, normally rules out univariate analysis techniques, which is why multivariate analysis methods were employed to elucidate the effects of the target metabolites, glucose and lactate (Rathore *et al.*, 2015, Sokolov *et al.*, 2015). Analysing the data was carried out using a combination of DoEMan

and PCA (2.3 & 2.4) to determine both the optimum pre-processing techniques and model calibration/validation data sets, to produce the final refined metabolite models.

4.1.1 Aim & Objectives

The aim of this study was to initially characterise the CHO cell culture, at a low passage number, in the commercial CD CHO medium in the batch/fed batch systems and to investigate the combined use of at-line NIR, MIR and Raman spectroscopy to monitor key metabolites in a number of CHO 42 bioprocesses, with the objective of developing metabolite models to aid in advancing towards real time monitoring. This spectral data would also be used in determining the feasibility of data fusion, to produce a single 'fused' model that in theory should have reduced errors of prediction than the individual models for each technique.

As these processes were to form the baseline for all the future work it was essential to fully characterise the CD CHO medium used. The changing nature of these types of process-fluids, caused by varying concentrations of cellular material and other components, means that application of spectroscopic techniques is not straightforward. However a number of studies have reported the success of NIR, MIR and Raman, either individually or in combination, for process monitoring of CHO cell cultures which is why these techniques were initially selected for the metabolite modelling (Hakemeyer *et al.*, 2012, Whelan *et al.*, 2012, Clavaud *et al.*, 2013, Sandor *et al.*, 2013, Pais *et al.*, 2014). These authors discuss the approaches they took to apply the respective spectroscopic techniques to CHO culture bioprocesses. They demonstrate that while models for some of the target metabolites/products were constructed the main challenges in the construction were the selection of calibration/validation datasets, separation of the target metabolites/products and processing of data.

In a previous study glucose/lactate pure component and media samples were analysed using the at-line instruments, to determine the appropriate spectral regions to focus on during subsequent samples spectral processing. These spectral regions were selected based on both the observed activity and that which has been reported in the literature (Williams and Fleming, 1995, Colthup, 2012, Nyquist and Kagel, 2012, Smith and Dent, 2013). In the present study the suitability of the at-line NIR, MIR and Raman spectrometers would also be assessed at this point and the instruments would be optimised for data collection. Utilising collected spectral data and the off-line reference data from the bioprocess samples, quantitative models would be produced for the most prominent metabolites, glucose and lactate. These models would undergo a validation using selected samples from among the replicate batches, which were not incorporated into the model calibration.

Although this research was carried out at-line, it creates a solid foundation on which to develop into processes with more complex media and could also facilitate a transition into developing an on-line process monitoring system.

4.1.2 Novelty

Although some research has demonstrated the effects of high passage number and chemically defined media upon CHO cultures, relating to protein expression or cell viability (Beckmann *et al.*, 2012, Veith *et al.*, 2016), there has been relatively little focus on lower passage number cultures and their media requirements. In terms of early culture development this offers the opportunity, through characterisation of low passage cultures in chemically defined media, to create new tailored media for early stages of bioprocesses development.

While a number of studies over the past few years have focused on the various capabilities NIR, MIR and Raman spectroscopies in biopharmaceutical applications, to the author's knowledge none have taken a DoE approach to data pre-processing and few have systematically investigated all three of the outlined techniques to the same process.

It is well documented that spectra are subject to numerous variations which make spectral interpretation difficult. This is where application of pre-processing techniques and PCA are introduced to enhance the spectra and allow the relevant information to be elucidated (Lourenco *et al.*, 2012). However, selection of pre-processing methods is often based on an *a priori* approach utilising previously reported studies or a desired effect (i.e. removal of background noise or baseline correction), rather than a systematic DoE approach which also incorporates previous knowledge of the matrix.

Even though there have been a number of studies which have utilised either one or two of the aforementioned spectroscopic techniques for biopharmaceutical process monitoring (Hakemeyer *et al.*, 2012, Ashton *et al.*, 2013, Sandor *et al.*, 2013) to the author's knowledge there are only a limited number of studies which have used all three. The potential advantage of utilising the three techniques is that the overlapping data sets could provide far more accurate process information.

Data fusion has been shown to be an effective method of combing the data for multiple compatible techniques, across a range of industries and processes, to produce models with lower errors of prediction than the individual techniques (Dearing *et al.*, 2011, Bachmann *et al.*, 2013, Chen *et al.*, 2015). However application of this strategy in the biopharmaceutical has been extremely limited and therefore offers a unique opportunity for improving process monitoring and modelling.

4.2 Results

4.2.1 Spectral Region Assignment

For both the batch and fed batch processes the initial spectral regions assigned to the two metabolites of interest, glucose and lactate, were based upon both previous work (data not shown) and those regions reported in the literature.

There are a number of regions where the signals representative of the glucose and lactate appear in the NIR spectrum, but only the combination band region (2200-2400 nm), which represents stretching and bending vibrations of C-H bonds, was selected for this research. This was for two reasons, it covered the broadest wavelengths of the spectrum that encompass the various C-H bonds found in the target metabolites and more importantly has been previously shown to be better for modelling (Chen *et al.*, 2004, Goodarzi *et al.*, 2015).

For the MIR spectrum the region between 900-1200 cm^{-1} was selected due to the series of distinct peaks reported to be present, which correspond to the target metabolites (Table 4.1) (Rhiel *et al.*, 2002, Lu *et al.*, 2011).

Table 4.1: Specific peaks of target metabolites.

Metabolite	Peak Region (cm^{-1})	Spectral Feature
Glucose	1150, 1106, 1079 & 1034	Vibrational C-O-C/C-O-H & Stretching C-O
Lactate	1123 & 1040	Carboxyl & Primary Alcohols

The region between, 800-1200 cm^{-1} was focused on for the Raman analysis of the process fluids. The particular peaks focused on were at 1072 cm^{-1} and 1128 cm^{-1} for

glucose, which corresponded to the stretching of the C-O and the C-O-H bending respectively. For lactate the $C\text{-}COO^-$ stretching peak at 855 cm^{-1} was utilised (Mehdizadeh *et al.*, 2015, Singh *et al.*, 2015).

4.2.2 CD CHO Batch Process Analysis

4.2.2.1 Reference Analysis

Initially the CHO 42 batch processes in CD CHO chemically defined medium were investigated to form a baseline, by characterising the cell metabolism (glucose utilisation/lactate production), live cell density/viability (Figure 4.1) and hIgG production. The highest observed average cell concentration was on day 7 ($5.8x10^6$ cells/mL), also coinciding with the first observed decrease in cell viability and increased lactate concentration. However despite the increased density, the glucose concentration was already depleted to nearly 1 g/L. The subsequent glucose deficiency and lactate accumulation led to the cultures being terminated on day 13, as it resulted in the average live cell viability reaching 0%. The hIgG production has been normalised as a percentage of the highest concentration achieved during the process due to commercial sensitivity, however it can be noted that the increase in concentration coincides with the increase in cell density. The subsequent decrease at the end of the process could be the result of degradation of the hIgG. See appendix Table A.1 for the corresponding values to Figure 4.1 and appendix Figure A.1 the bioreactor process condition values (pH, dO_2, Agitation & Temperature).

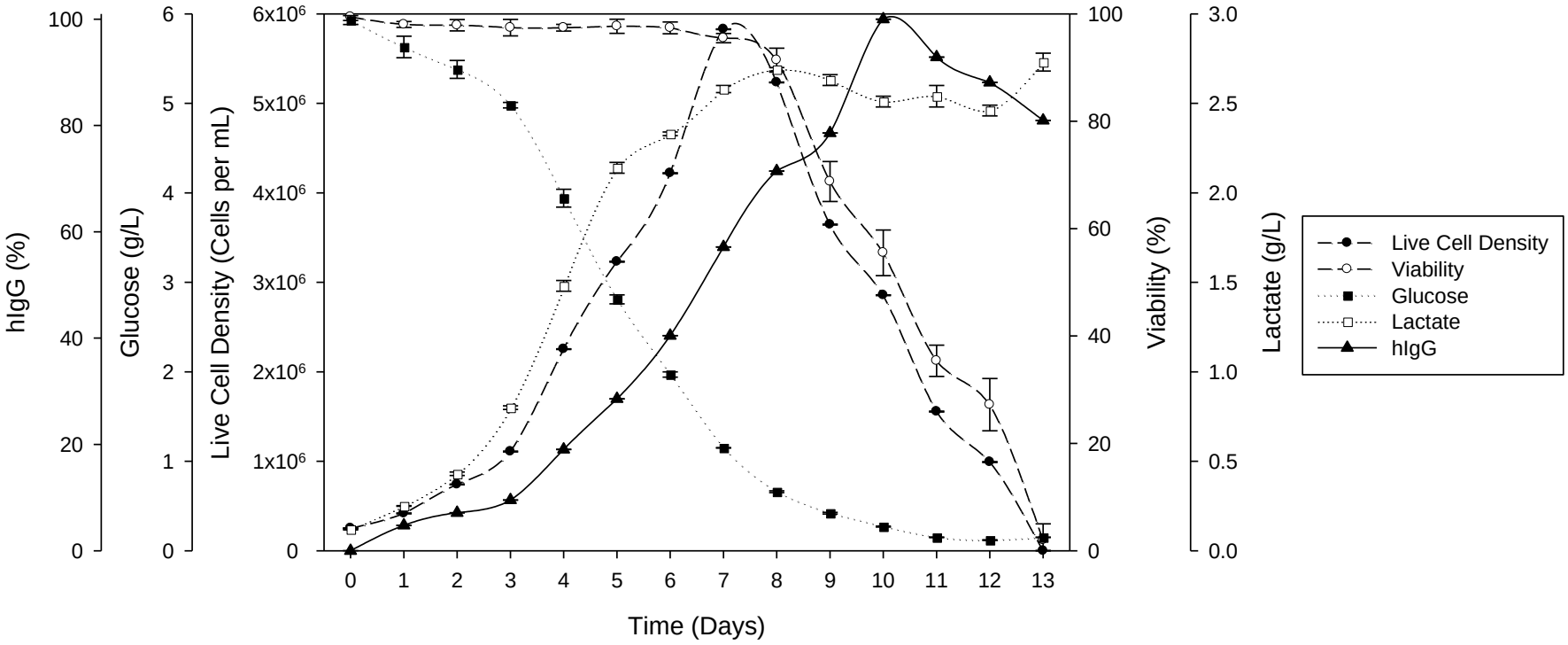


Figure 4.1: Time course of process data from CHO 42 batch culture in CD CHO medium [average of triplicate measurements from three bioreactors plotted with standard deviation].

4.2.2.2 Raw Process Spectra

As shown by the raw process spectra from the NIR and MIR (Figure 4.2 & 4.3) there are no clear discernible signals due to the intensity of the water absorbance, which is consistent with the literature (Beutel and Henkel, 2011, Reich, 2016).

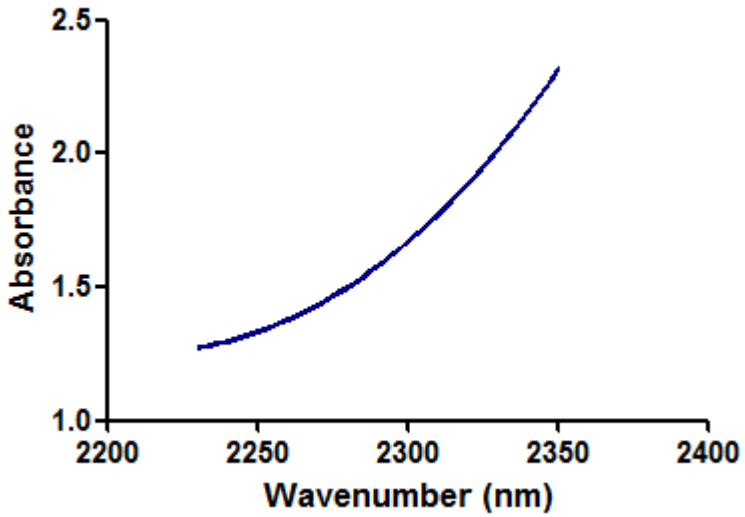


Figure 4.2: Batch CD CHO NIR absorbance of combination band region (stretching and bending vibrations for C–H) from all processes [Colour Code Light to Dark = Bioreactor 1 to 3] (Processing: Average of triplicate measurements from each bioreactor).

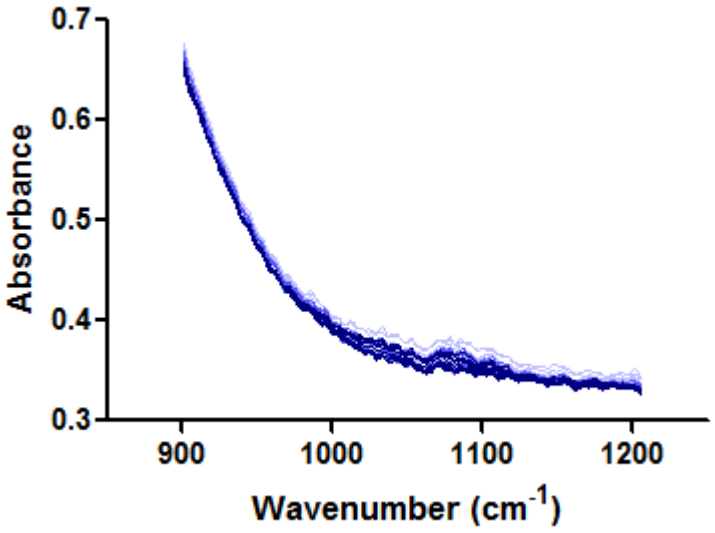


Figure 4.3: Batch CD CHO MIR absorbance of 900-1200 cm^{-1} region from all processes [Colour Code Light to Dark = Bioreactor 1 to 3] (Processing: Average of triplicate measurements from each bioreactor).

However, the Raman spectra displayed activity at only two of the specific peaks, 855 cm^{-1} and 1128 cm^{-1}, in the region identified from the literature (Mehdizadeh *et al.*, 2015, Singh *et al.*, 2015) which was why the area around these peaks, 835-865 cm^{-1} and 1100-1150 cm^{-1}, were selected for all future analysis (Figure 4.4).

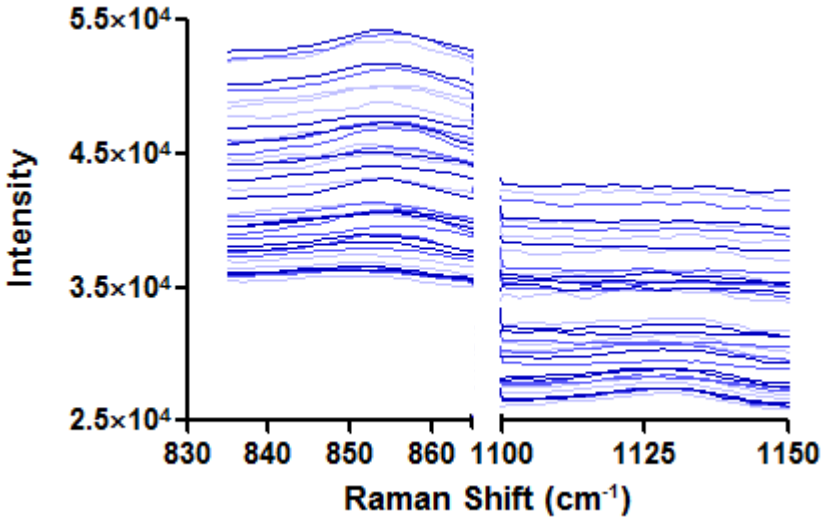


Figure 4.4: Batch CD CHO Raman scattering of 835-865 cm^{-1} and 1100-1150 cm^{-1} regions from all processes [Colour Code Light to Dark = Bioreactor 1 to 3] (Processing: Average of triplicate measurements from each bioreactor).

4.2.2.3 DoEman

As previously stated a combination of using *a priori* knowledge from the literature and a systematic DoE approach was taken to optimise the pre-processing of the raw spectral data and selection of the calibration/validation datasets, before constructing the models. Initially, a DoE approach was taken using the DoEman software outlined in 3.8.2, to determine what the optimal pre-processing techniques were. A number of commonly reported techniques from the literature, known to be useful in spectral pre-processing, were selected for investigation.

As the spectral data require a series of pre-processing treatments to be investigated, it was essential that the treatments were applied in the correct sequence to avoid any negative influences either upon or between the treatments (Flåten and

Walmsley, 2003, Gerretzen *et al.*, 2015). The experimental design utilised in DoEman analysis of the CD CHO batch cultures in Table 3.2.

The first variable applied was the regression method, which compared PLS1 to PLS2. As Savitzky-Golay has been reported in either first or second derivative, it had to be applied in three levels within DoEman. The filter width and polynomial order for the first and second derivation was set to, 21 and 2, respectively. The transformative techniques MSC, which deals with light scattering effects, and SNV for spectral averaging were also investigated, because of their ability to correct spectral data for scaling and offset (baseline) effects. Next to be investigated was one of the most common pre-processing techniques, mean centering of the data and finally the number of latent variables was investigated which ranged from one to ten.

The NIR results from the DoEman plots for both the glucose and lactate indicated that the following combination of pre-processing techniques, PLS1, 2nd Derivative (21, 2), MSC and Mean Centering, would yield the most optimal results from the data, when constructing the PLS models. The optimal MIR data pre-processing regime for glucose and lactate was identified by the DoEman plots as a combination of MSC and mean centering, which were used for the PCA to determine the separation of calibration/validation datasets. Similar to the MIR results the DoEman plots for the Raman glucose and lactate data specified that the optimal pre-processing techniques were, MSC and mean centering, which were subsequently applied in the PCA to determine the calibration/validation datasets.

4.2.2.4 PCA Scores/Plots

The next step in the spectral data optimisation was to select calibration and validation datasets. For this PCA was carried out to ensure that the spectral variation between the

three batches was captured. The pre-processing techniques indicated by the DoE to be the most optimal were applied to the data in the PCA, as was an internal validation (Table 4.3).

Table 4.3: Spectral pre-processing PCA parameters and statistical values

Technique	Pre-processing	Cross Validation	No. of Latent Variables	Variance Captured
NIR	2^{nd} Derivative, MSC & Mean Centering	Venetian Blinds	6	90.89 %
MIR	MSC & Mean Centering	Venetian Blinds	6	70.47 %
Raman	MSC & Mean Centering	Venetian Blinds	6	88.78 %

Using the previously outlined parameters for the NIR PCA (Table 4.3), a trend was observed within the PCA scores [PC1 vs PC2] (Figure 4.5) as all three batches seem to overlap, indicating similarity between the batches and samples. Early process samples correlate to PC1 and over time as the process continues the samples begin to shift and correlate to PC2. Based on these observations, batches 2 and 3 were selected as the calibration dataset which would be used to predict the validation dataset, batch 1.

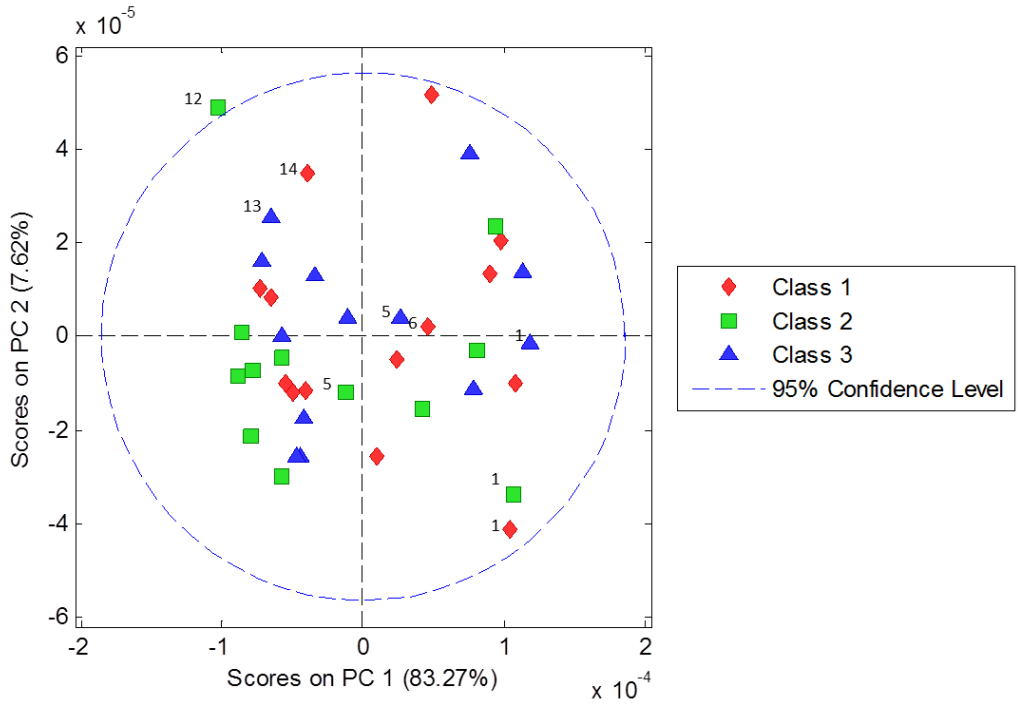


Figure 4.5: Batch CD CHO NIR PCA scores of the combination band region in all processes, with process sample numbers represented [Class=Bioreactor 1-3] (Processing: Savitzky-Golay Second Derivative; MSC; Mean Centered; Average of triplicate measurements per batch).

The MIR PCA, using the parameters from Table 4.3, resulted in the PCA scores [PC1 vs PC2] plot displayed in Figure 4.6. While the results from the batches would be expected to overlap the three batches seem to be individually distributed, indicating differences between the batches. But there is some overlap between batches 2 and 3, suggesting similarities in these samples. There does seem to be a trend within each batch that is similar, with majority of early process samples correlating to PC1 and later process samples to PC2. From these observations, batches 1 and 3 were assigned to the calibration dataset to predict the validation dataset, batch 2.

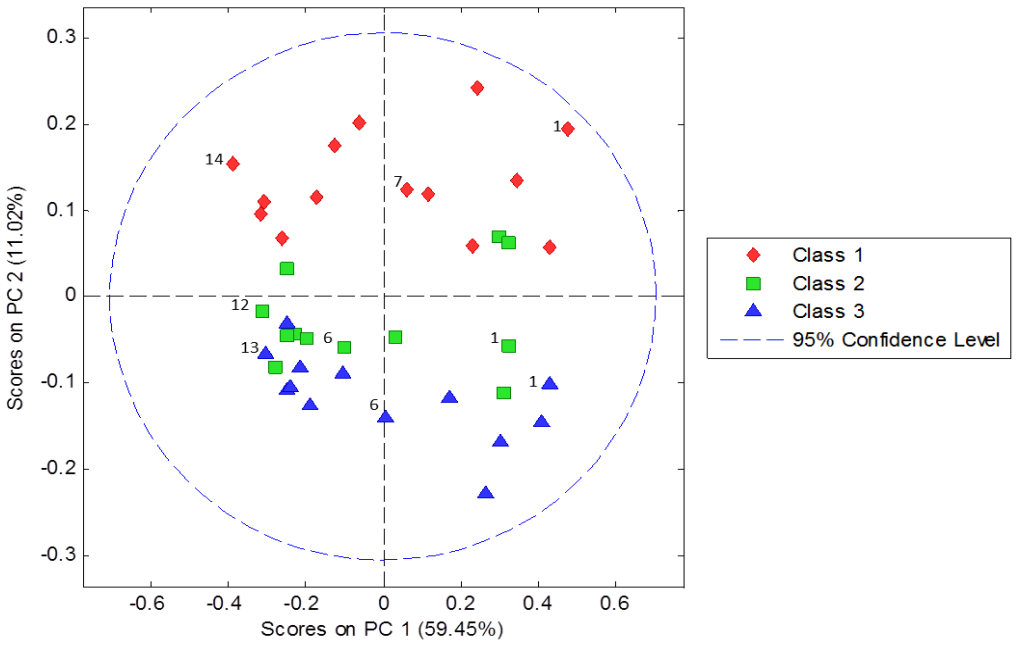


Figure 4.6: Batch CD CHO MIR PCA scores of the 900-1200 cm^{-1} region in all processes, with process sample numbers represented [Class=Bioreactor 1-3] (Processing: MSC and Mean Centered; Average of triplicate measurements per batch).

Similar to the NIR results the Raman PCA (Figure 4.7) display an apparent trend, as all three batches are overlapping, indicating similarity and reproducibility. Earlier process samples correlate to PC1 and then develop a correlation with PC2, as the process continues. From these observations, batches 1 and 2 were selected as the calibration dataset which would be used to predict the validation dataset, batch 3.

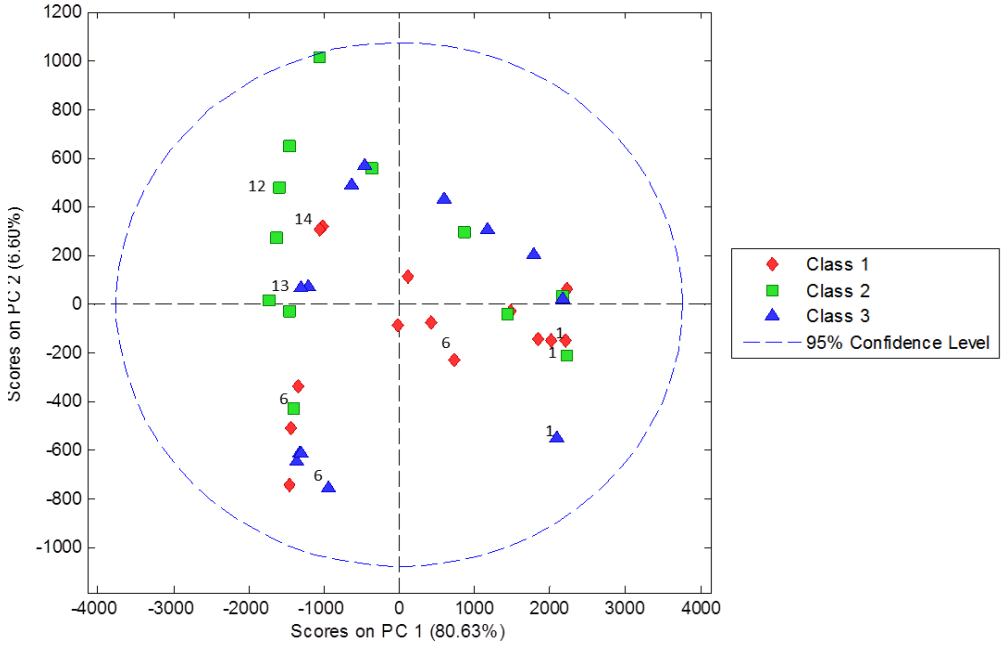


Figure 4.7: Batch CD CHO Raman PCA scores of the 835-865 cm^{-1} and 1100-1150 cm^{-1} regions from all processes, with process sample numbers represented [Class=Bioreactor 1-3] (Processing: MSC and Mean Centered; Average of triplicate measurements per batch).

Although it is expected for the batches to overlap in the PCA scores plots, there is enough overlap between the batches to indicate similarity despite some deviation. From the observed trends, move from PC1 to PC2 as time passes, within the PCA scores plots of each technique it can be inferred that PC1 and PC2 correlate to glucose and lactate, respectively.

4.2.2.5 Processed Spectra

Once the selected pre-processing parameters from the DoEman are applied to the NIR spectral data, distinct peaks and troughs can be observed which correspond to the absorbance intensity within the CH region (Figure 4.8). The changes in height and depth of

the peaks follow the concentration of glucose/lactate trends within the reference data from all batches.

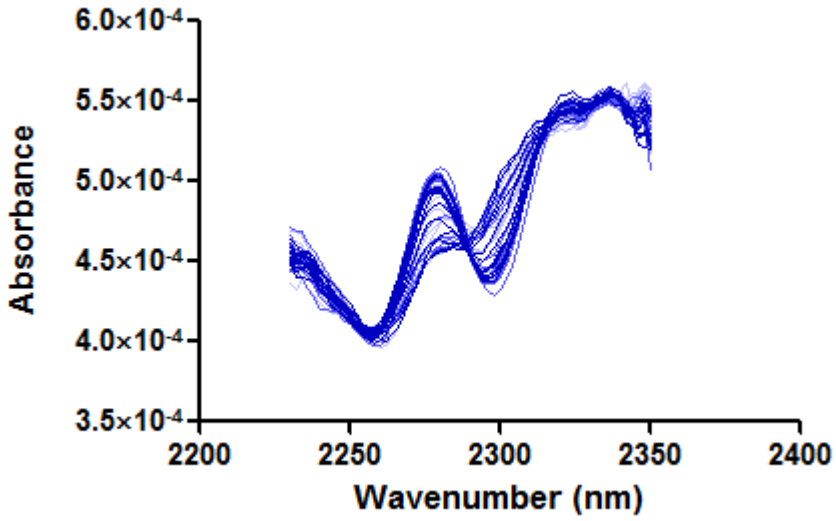


Figure 4.8: Preprocessed batch CD CHO NIR absorbance of CH region of all processes [Colour Code Light to Dark = Bioreactor 1 to 3] (Processing: Savitzky-Golay Second Derivative; MSC; Mean Centered; Average of triplicate measurements from each bioreactor).

Despite application of the optimised pre-processing techniques, there do not appear to be any distinct signals corresponding to the peak areas of interest within the MIR data (Figure 4.9). This lack of activity indicates that the PLS modelling might not be able to predict either glucose or lactate.

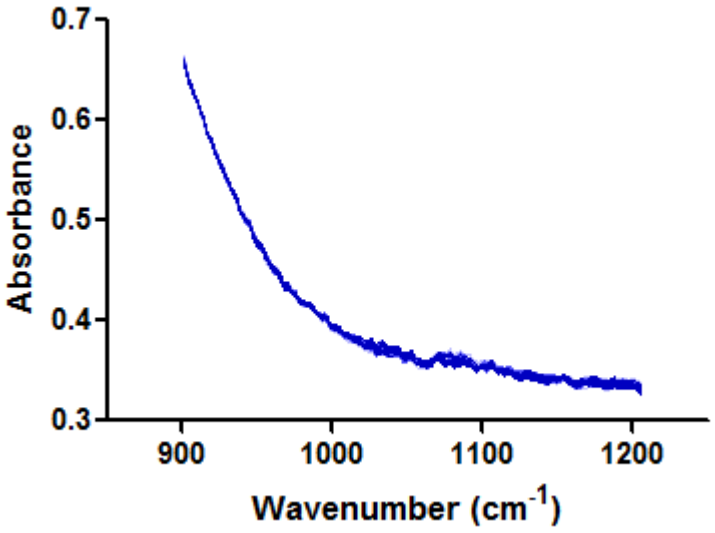


Figure 4.9: Preprocessed batch CD CHO MIR absorbance of 900-1200 cm^{-1} region from all processes [Colour Code Light to Dark = Bioreactor 1 to 3] (Processing: MSC; Mean Centered; Average of triplicate measurements from each bioreactor).

Like the NIR data, after the application of the pre-processing treatments the target peaks within the Raman data are elucidated (Figure 4.10). Again peak intensity seems to correspond to the concentration of glucose and lactate within the samples of all three batches.

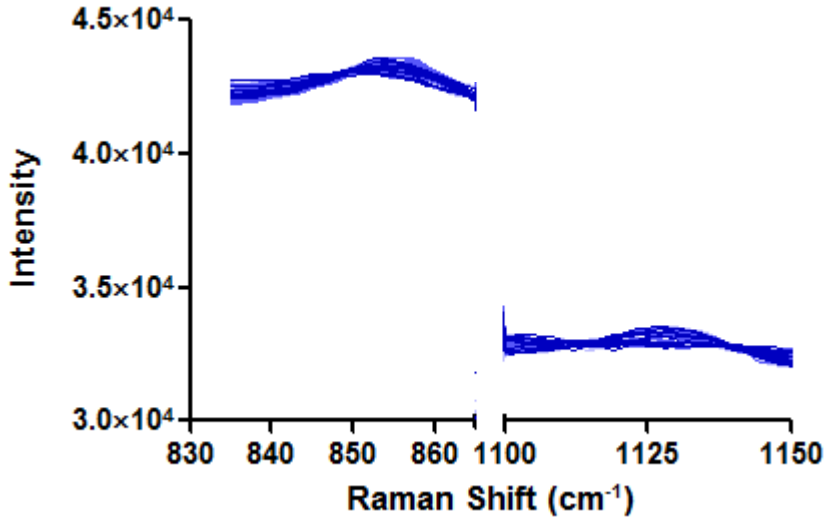


Figure 4.10: Processed batch CD CHO Raman scattering of 835-865 cm^{-1} and 1100-1150 cm^{-1} regions from all processes [Colour Code Light to Dark = Bioreactor 1 to 3] (Processing: MSC; Mean Centered; Average of triplicate measurements from each bioreactor).

4.2.2.6 Partial Least Squares (Calibration & Validation)

Once the spectral data optimisation was completed and the calibration and validation datasets selected, the final stage was to carry out the PLS modelling. The PLS models and their Root Mean Square Errors (RMSE) produced from the NIR data indicate, good model fit and low errors of prediction for both glucose (Figure 4.11) and lactate (Figure 4.12). Although there is some shift in the lactate model, possibly due to selection of samples in the prediction dataset.

[A]

	Calibration	Internal Validation	External Validation (Prediction)
RMSE (g/L)	0.30	0.70	0.45
R^2	0.98	0.91	0.96

[B] **[C]**

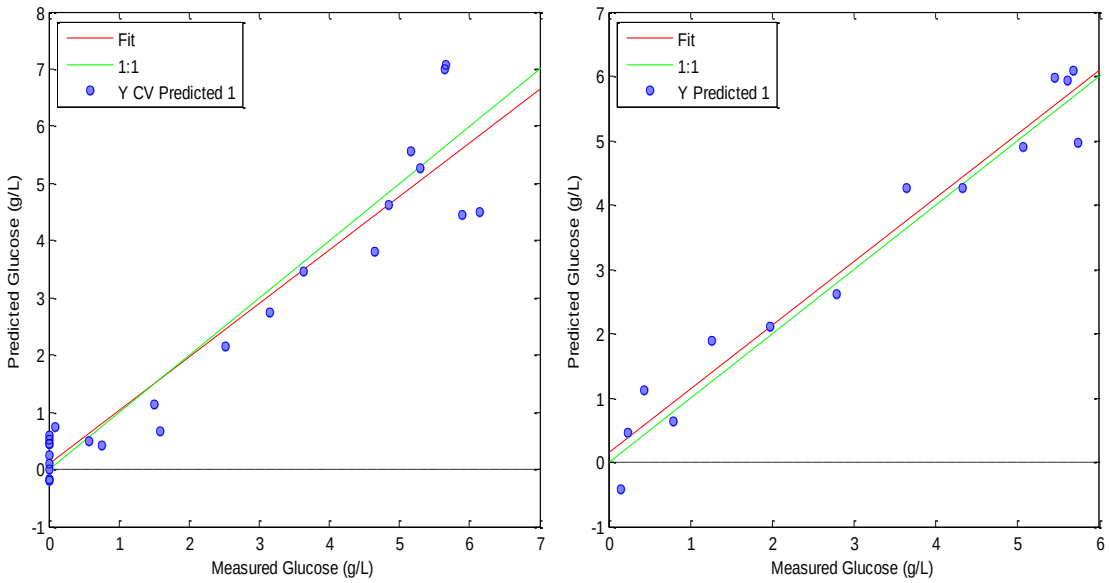


Figure 4.11: NIR batch CD CHO glucose PLS [A] RMSE & R^2 values [B] Calibration & Internal Validation model [C] External Validation (Prediction) model; with the selected samples (blue points), the theoretical 'perfect' fit (green line) and the actual fit (red line).

[A]

	Calibration	Internal Validation	External Validation (Prediction)
RMSE (g/L)	0.17	0.37	0.45
R^2	0.97	0.91	0.94

[B] **[C]**

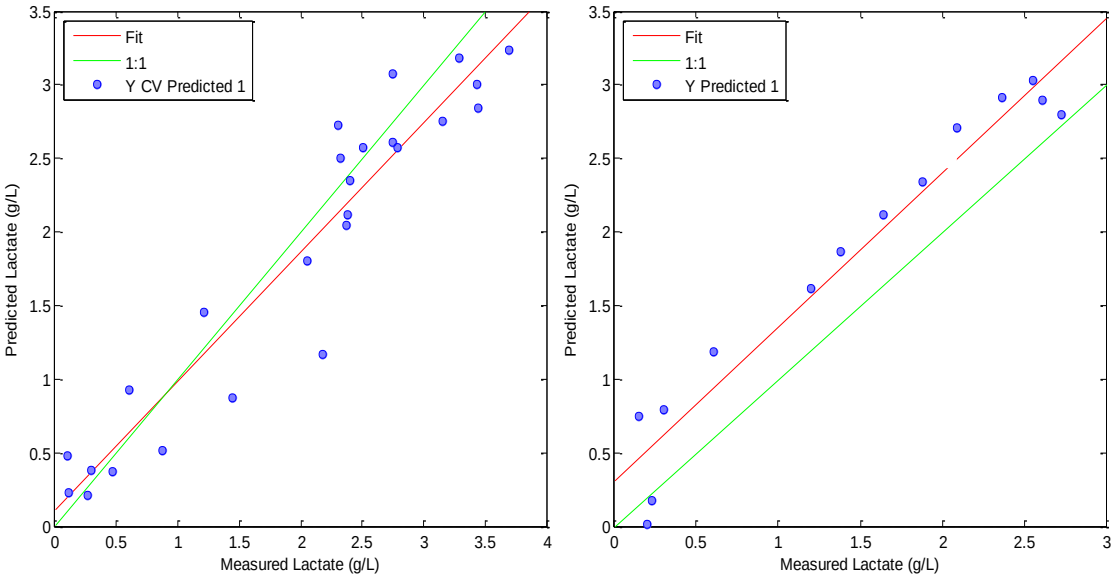


Figure 4.12: NIR batch CD CHO lactate PLS [A] RMSE & R^2 values [B] Calibration & Internal Validation model [C] External Validation (Prediction) model; with the selected samples (blue points), the theoretical 'perfect' fit (green line) and the actual fit (red line).

Unlike the NIR PLS models, the MIR PLS models and their RMSE indicate a poor model fit and high errors of prediction for both glucose (Figure 4.13) and lactate (Figure 4.14). As the batches display little similarity in the PCA plot, it is not unexpected that the model would not predict well due to the high level of variance.

[A]

	Calibration	Internal Validation	External Validation (Prediction)
RMSE (g/L)	0.33	2.77	2.79
R^2	0.98	0.01	0.00

[B] **[C]**

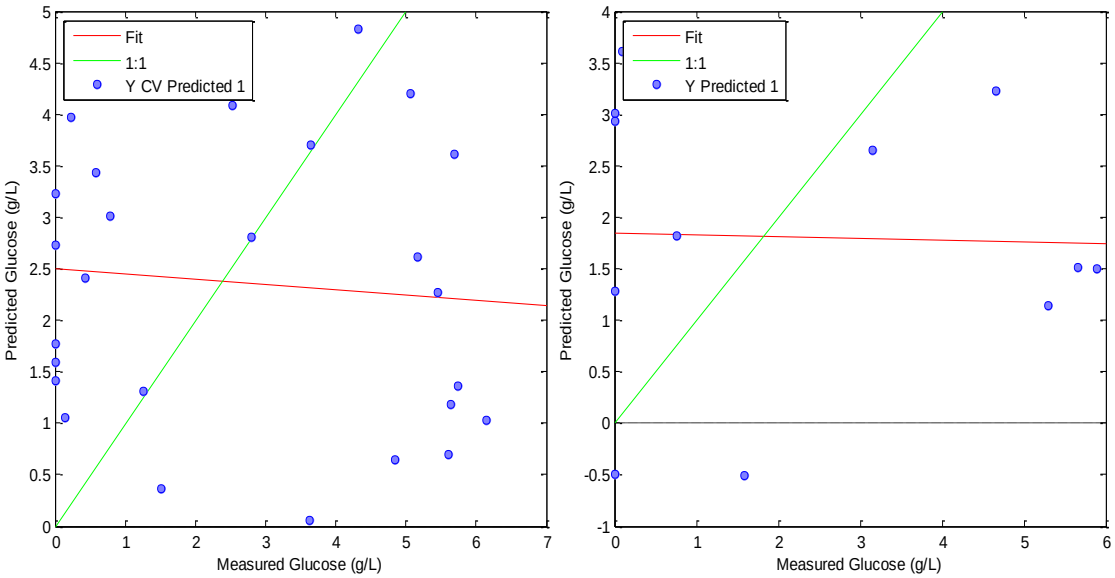


Figure 4.13: MIR batch CD CHO glucose PLS [A] RMSE & R^2 values [B] Calibration & Internal Validation model [C] External Validation (Prediction) model; with the selected samples (blue points), the theoretical 'perfect' fit (green line) and the actual fit (red line).

[A]

	Calibration	Internal Validation	External Validation (Prediction)
RMSE (g/L)	0.13	1.15	1.43
R^2	0.98	0.01	1.83

[B] **[C]**

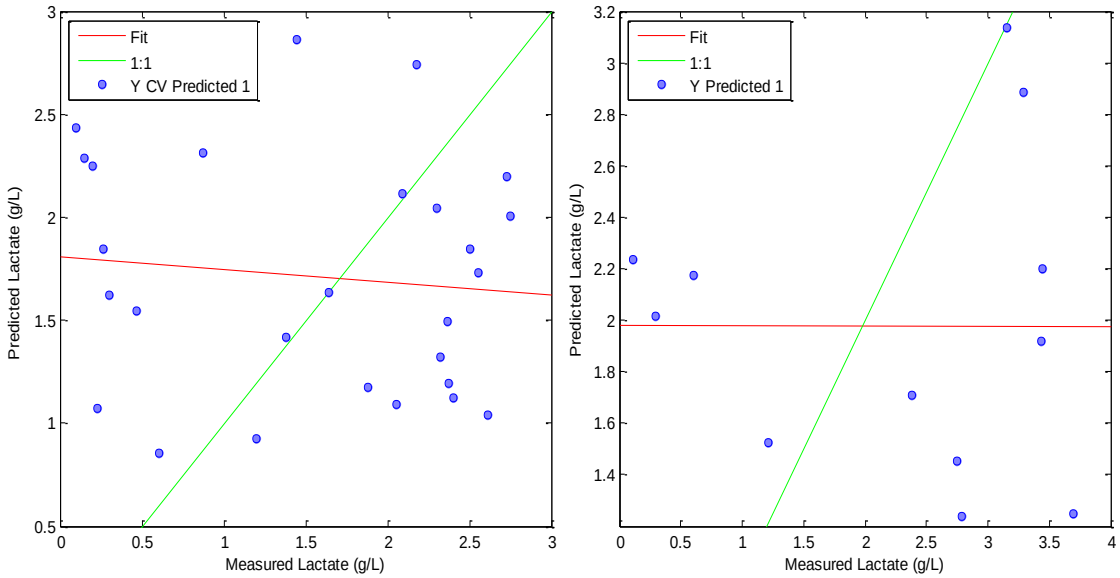


Figure 4.14: MIR batch CD CHO lactate PLS [A] RMSE & R^2 values [B] Calibration & Internal Validation model [C] External Validation (Prediction) model; with the selected samples (blue points), the theoretical 'perfect' fit (green line) and the actual fit (red line).

Similar to the MIR PLS models, the Raman PLS models and their RMSE indicate a poor model fit and high errors of prediction for both glucose (Figure 4.15) and lactate (Figure 4.16). Despite the batches displaying similarity in the PCA plot, it is unexpected that the model would not predict. However there may also be unseen effects from other components with the sample matrix or poor selection of calibration/validation datasets, which have resulted in poor model performance.

[A]

	Calibration	Internal Validation	External Validation (Prediction)
RMSE (g/L)	0.84	2.58	3.11
R^2	0.87	0.24	0.43

[B] **[C]**

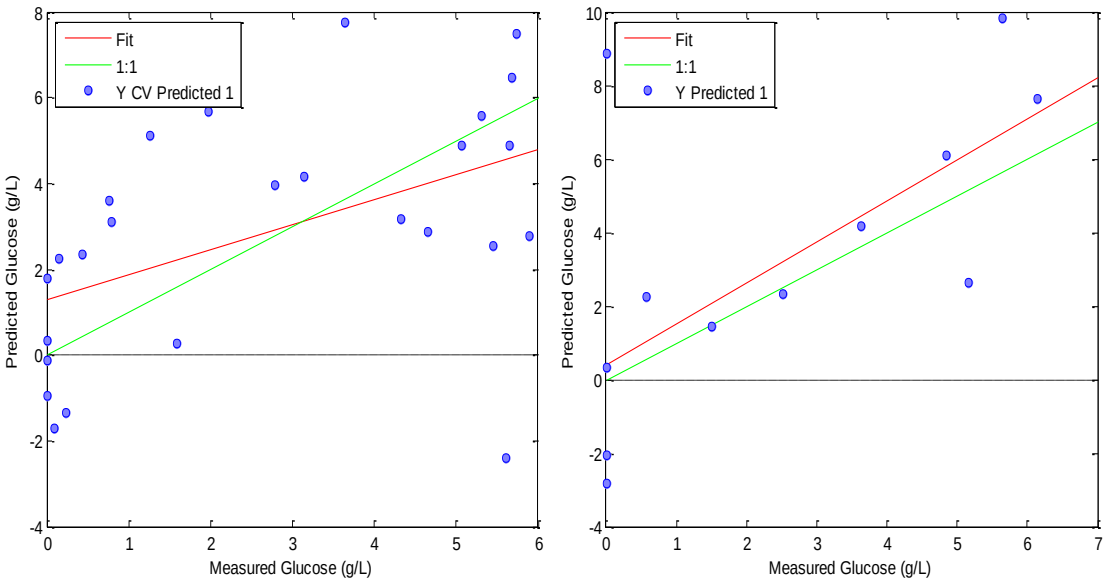


Figure 4.15: Raman batch CD CHO glucose PLS [A] RMSE & R^2 values [B] Calibration & Internal Validation model [C] External Validation (Prediction) model; with the selected samples (blue points), the theoretical 'perfect' fit (green line) and the actual fit (red line).

[A]

	Calibration	Internal Validation	External Validation (Prediction)
RMSE (g/L)	0.43	1.38	1.62
R^2	0.87	0.29	0.51

[B] **[C]**

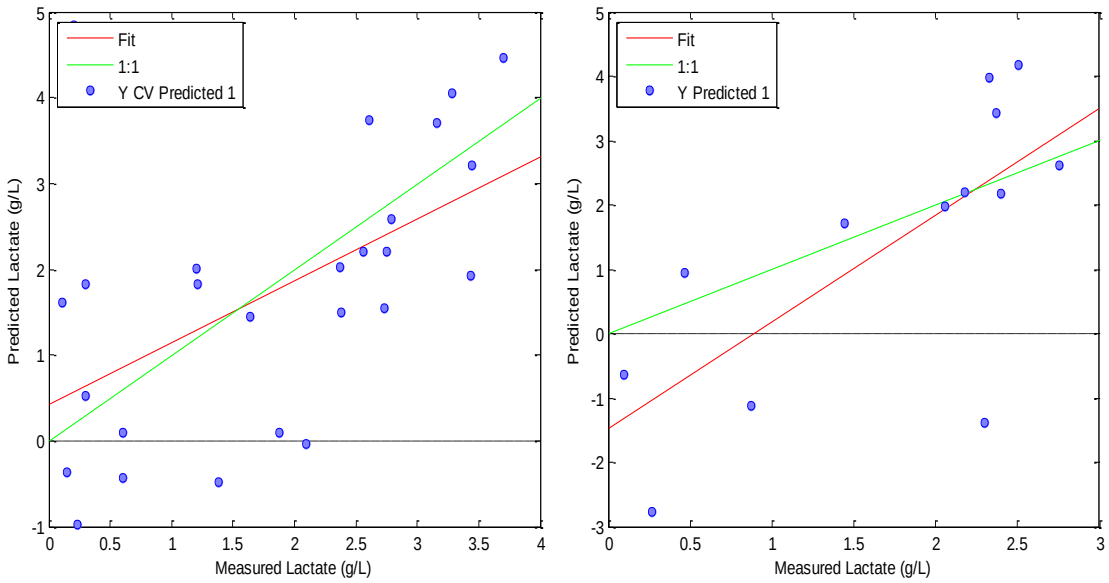


Figure 4.16: Raman batch CD CHO lactate PLS [A] RMSE & R^2 values [B] Calibration & Internal Validation model [C] External Validation (Prediction) model; with the selected samples (blue points), the theoretical 'perfect' fit (green line) and the actual fit (red line).

Overall the PLS modelling has produced results of varying success. The models generated from the NIR indicate suitability for predicting glucose and lactate. While the MIR and Raman models indicate poor suitability for predicting both metabolites.

4.2.3 CD CHO Fed Batch Process Analysis

4.2.3.1 Reference Analysis

Following the batch processes, a fed batch process in the CD CHO medium with addition of CHO CD EfficientFeed™ B (EFB) was carried out, for further process development. Using the

strategy outlined in Section 3.2.2.2, the processes characterised the cell metabolism (glucose utilisation/lactate production), live cell density/viability (Figure 4.17) and antibody titre. The highest observed average cell density was on day 7 ($5.2x10^6$ cells/mL) however this was lower than the batch processes. Despite the feed strategy, the process does not appear to have been affected in a positive manner. The overall lactate concentration was higher than in the batch process, it was also noted that it did decrease in the later stages of the processes. As with the batch process the fed batch culture was terminated on day 13, as the viability reach 0%. The hIgG production has been normalised as a percentage of the highest concentration achieved during the process due to commercial sensitivity, however it can be noted that the increase in concentration coincides with the increase in cell density. The subsequent decrease at the end of the process could be the result of degradation of the hIgG. See appendix Table A.2 for the corresponding values to Figure 4.17 and appendix Figure A.2 the bioreactor process condition values (pH, dO_2, Agitation & Temperature).

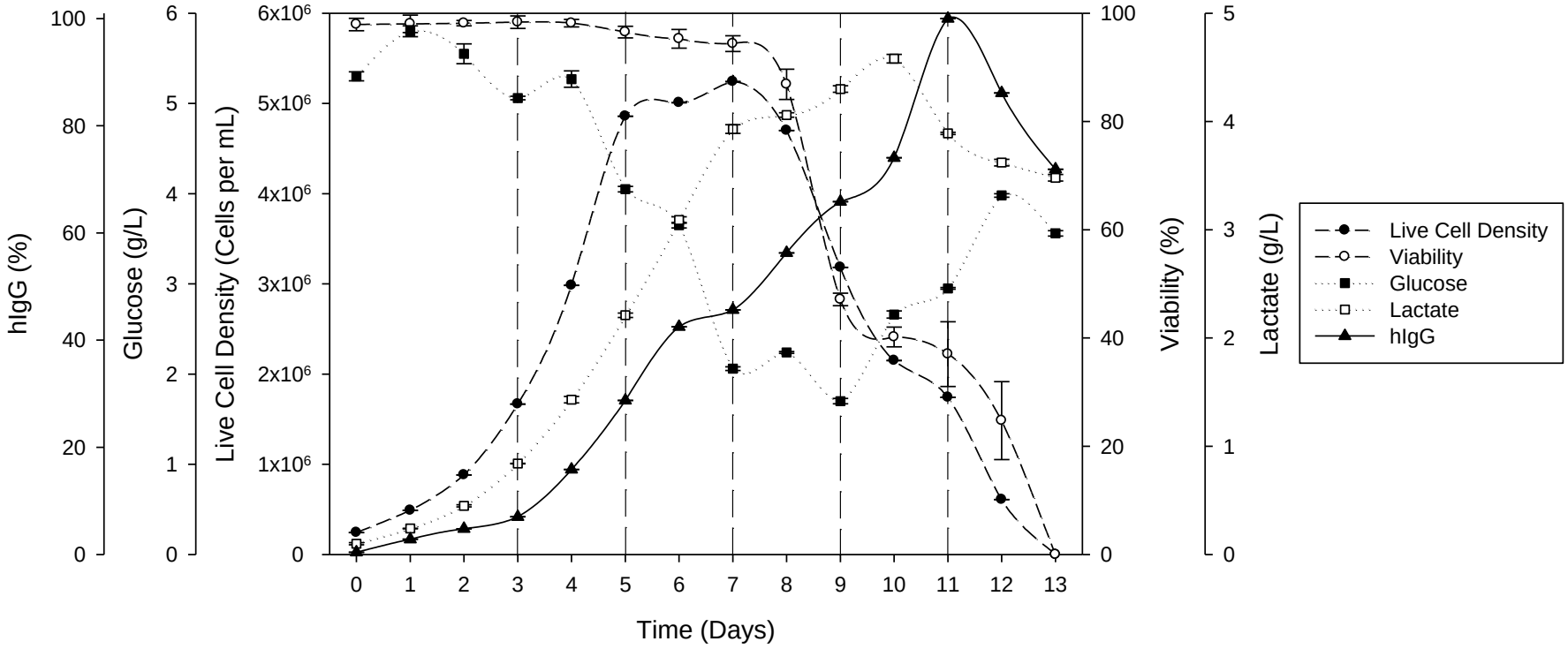


Figure 4.17: Time course of process data from CHO 42 fed batch culture in CD CHO medium. Vertical lines represent days when EfficientFeed™ B was fed to the culture [average of triplicate measurements from two bioreactors plotted with standard deviation error bars].

4.2.3.2 Raw Process Spectra

Similar to the batch processes, the raw process spectra from the NIR and MIR (Figure 4.18 & 4.19) do not exhibit any discernible signals due to the intensity of the water absorbance.

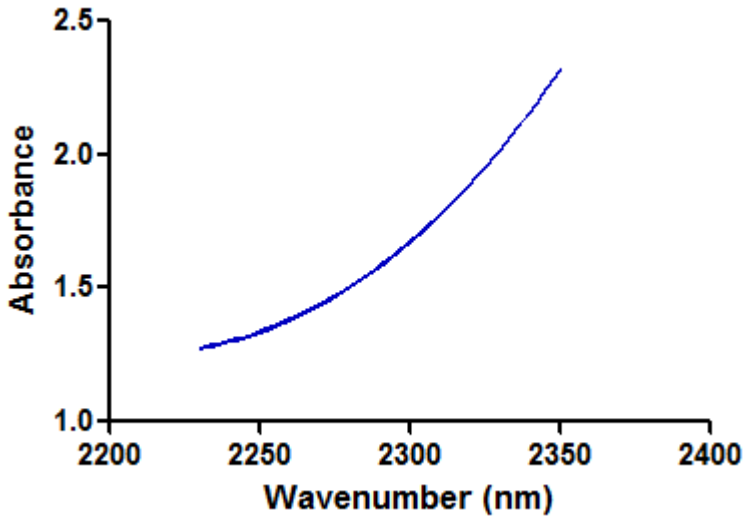


Figure 4.18: Fed batch CD CHO NIR absorbance of combination band region (stretching and bending vibrations for C–H) from all processes [Colour Code Light to Dark = Bioreactor 1 & 2] (Processing: Average of triplicate measurements from each bioreactor) Raw spectra MIR

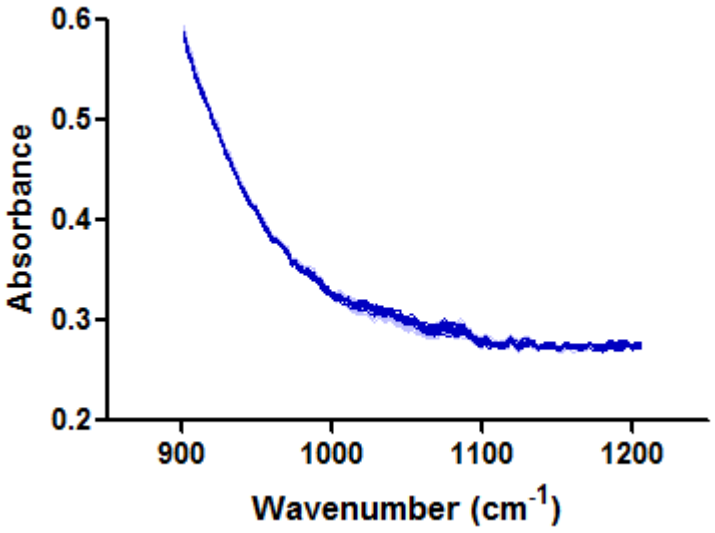


Figure 4.19: Fed batch CD CHO MIR absorbance of 900-1200 cm^{-1} region from all processes [Colour Code Light to Dark = Bioreactor 1 & 2] (Processing: Average of triplicate measurements from each bioreactor).

The Raman spectra again displayed activity at the 855 cm^{-1} and 1128 cm^{-1} peaks, however in comparison the peaks are less pronounced than in the batch cultures. For the PLS analysis the regions around the two peaks were selected, 835-865 cm^{-1} and 1100-1150 cm^{-1} (Figure 4.20).

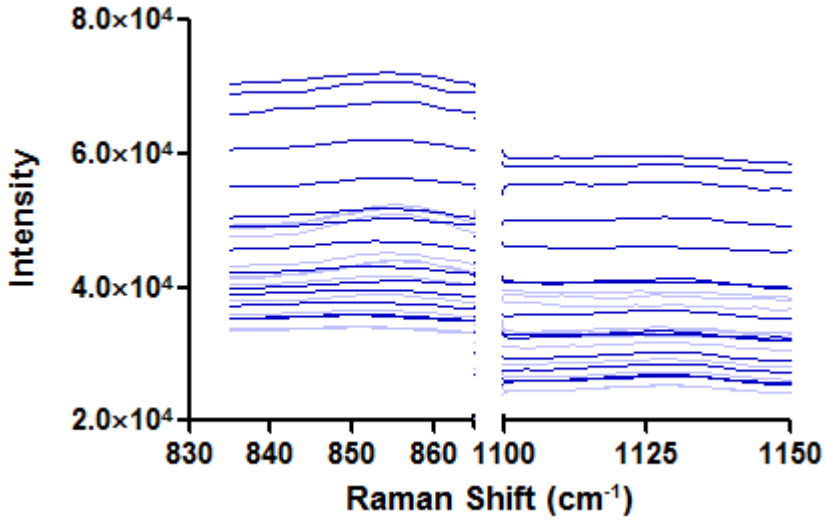


Figure 4.20: Fed batch CD CHO Raman scattering of 835-865 cm^{-1} and 1100-1150 cm^{-1} regions from all processes [Colour Code Light to Dark = Bioreactor 1 to 2] (Processing: Average of triplicate measurements from each bioreactor).

4.2.3.3 DoEman

The spectral pre-processing optimisation strategy outlined in Section 4.2.2.3 was also utilised in the analysis of the fed batch processes.

The results from the DoEman plots for each technique reflect the same outcomes which were observed from the analysis of the batch processes for both glucose and lactate. For the NIR data the optimal pre-processing strategy was indicated to be, 2nd Derivative (21, 2), MSC and Mean Centering. While for both MIR and Raman, MSC and mean centering were again identified as the most optimal for these datasets. These outcomes were then utilised in the PCA to determine the calibration/validation datasets.

4.2.3.4 PCA Scores/Plots

To select calibration and validation datasets, PCA was performed ensuring that the spectral variation between the two batches was captured. The pre-processing techniques indicated by the DoE to be the most optimal were applied to the data in the PCA, as was an internal validation (Table 4.4).

Table 4.4: Spectral pre-processing PCA parameters and statistical values

Technique	Pre-processing	Cross Validation	No. of Latent Variables	Variance Captured
NIR	2^{nd} Derivative, MSC & Mean Centering	Venetian Blinds	4	99.08 %
MIR	MSC & Mean Centering	Venetian Blinds	7	81.34 %
Raman	MSC & Mean Centering	Venetian Blinds	6	97.04 %

Using the previously outlined parameters for the NIR PCA (Table 4.4), there was a trend observed within the PCA scores [PC1 vs PC2] (Figure 4.21) of both batches. The earlier process samples correlated to PC1 and over time as the process continues the samples begin to shift and correlate to PC2. Based on these observations, two thirds of the data were selected from both batches to form the calibration data set, while the remaining third formed the validation dataset. However the batches themselves display limited overlapping indicating poor batch to batch differences.

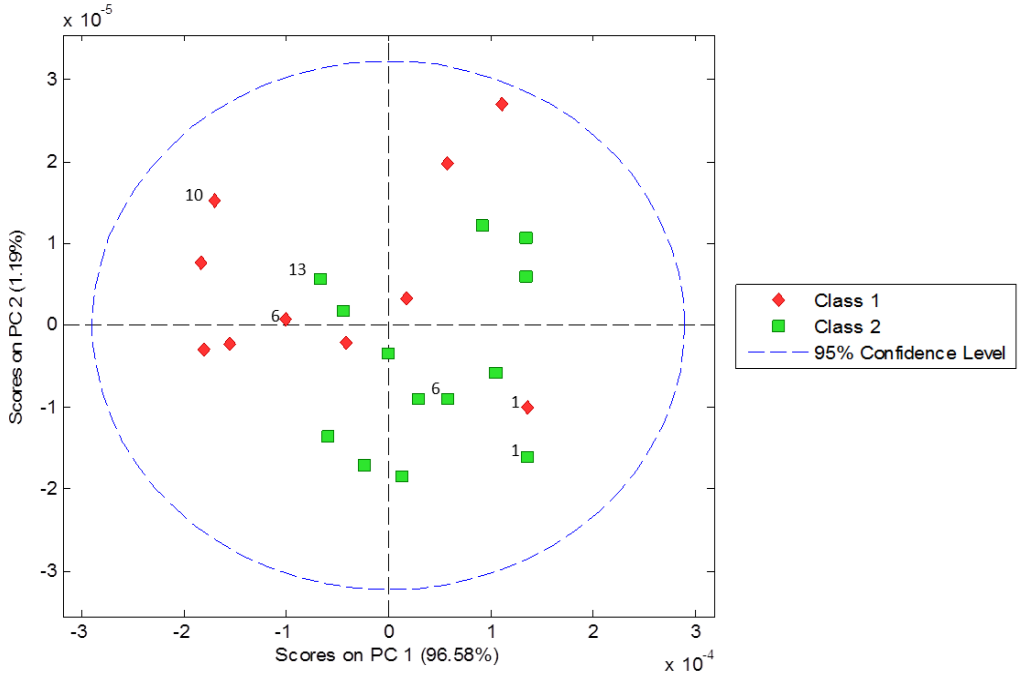


Figure 4.21: Fed batch CD CHO NIR PCA scores of the combination band region in all processes, with process sample numbers represented [Class=Bioreactor 1-2] (Processing: Savitzky-Golay Second Derivative; MSC; Mean Centered; Average of triplicate measurements per batch).

The MIR parameters from Table 4.4 when applied to the data, resulted in the PCA scores [PC1 vs PC2] plot displayed in Figure 4.22. There was no overlap within two batches, indicating no batch to batch similarity as they seem to be individually distributed. But there was again a trend within each batch which was similar to the previous processes, as the majority of early process samples correlated to PC1 and later process samples to PC2. Two thirds of the data were selected, which covered a wide distribution of the data, to form the calibration dataset and the remaining third was used for the validation dataset.

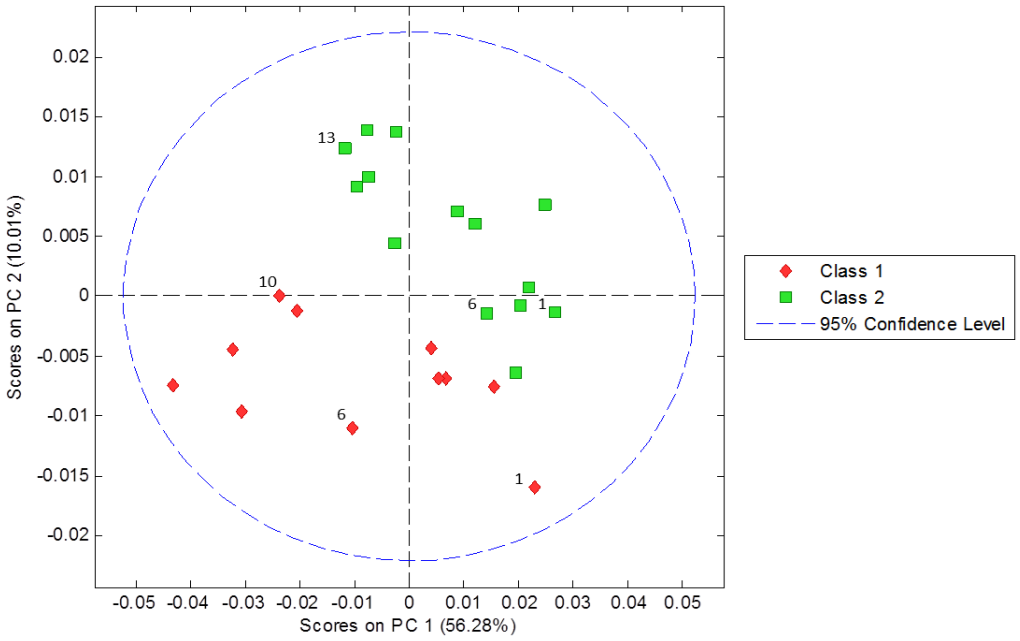


Figure 4.22: Fed batch CD CHO MIR PCA scores of the 900-1200 cm^{-1} region in all processes, with process sample numbers represented [Class=Bioreactor 1-2] (Processing: MSC and Mean Centered; Average of triplicate measurements per batch).

Results of the Raman PCA (Figure 4.23) display the opposite trend as the previous datasets where, earlier process samples correlate to PC2 and then develop a correlation with PC1, as the process continues. However there is only a minimal overlap between batches, but again from the data two thirds were selected as a calibration dataset and a third for the validation to be used in the PLS modelling.

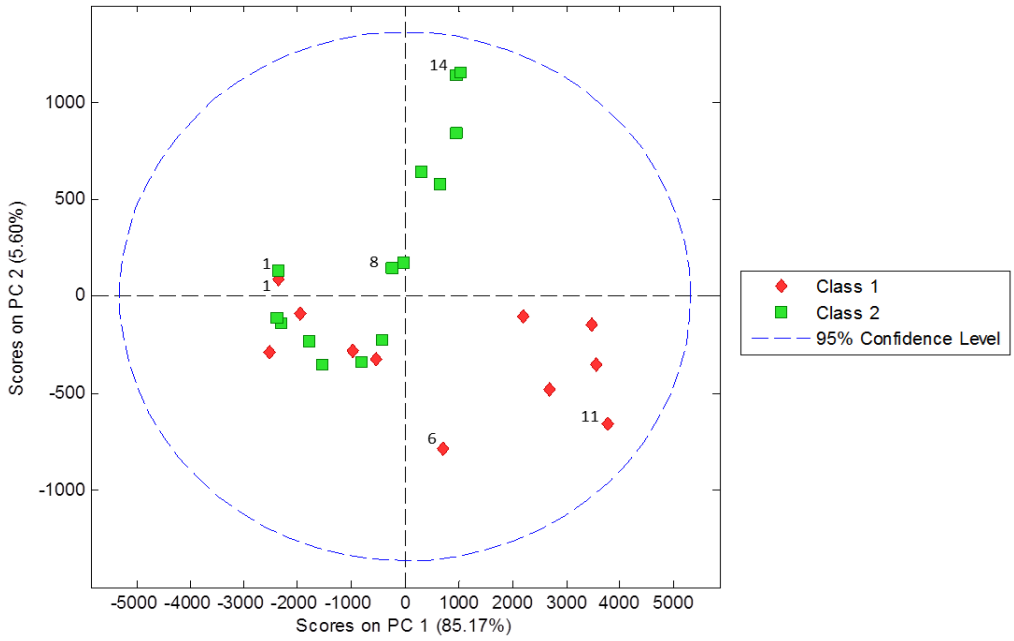


Figure 4.23: Fed batch CD CHO Raman PCA scores of the 835-865 cm^{-1} and 1100-1150 cm^{-1} regions from all processes, with process sample numbers represented [Class=Bioreactor 1-2] (Processing: MSC and Mean Centered; Average of triplicate measurements per batch).

Despite the absence of overlap in the fed batch PCA scores plots from the NIR and MIR, the presence of a trend from PC1 and PC2 again indicate a correlation to glucose and lactate, respectively. The alternate trend in the Raman data is unexpected and does indicated that modelling of this data would be challenging, due to the distribution of the samples.

4.2.3.5 Processed Spectra

Much like the batch data once the selected pre-processing parameters from the DoEman are applied to fed batch spectral data, distinct peaks and troughs can be observed which follow the concentration of glucose/lactate trends within the reference data from all

batches (Figure 4.24). However there is less overlap between the two batches which reflect the same trends in the PCA plots.

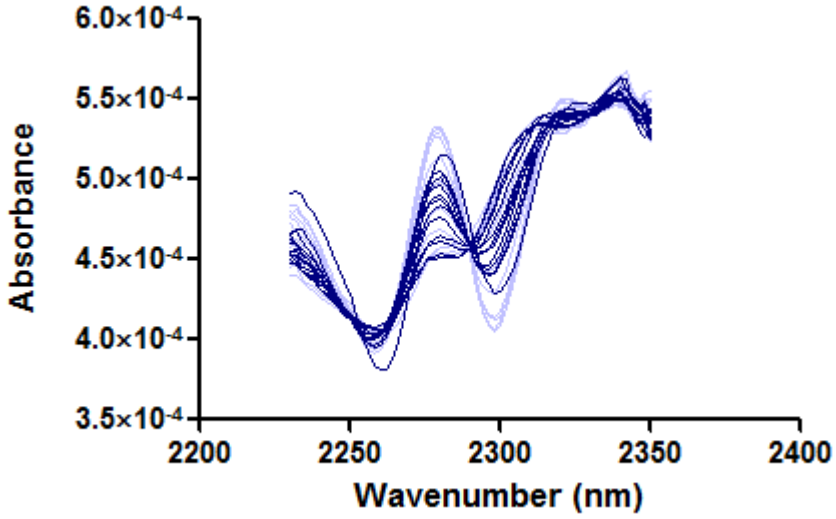


Figure 4.24: Preprocessed fed batch CD CHO NIR absorbance of CH region of all processes [Colour Code Light to Dark = Bioreactor 1 & 2] (Processing: Savitzky-Golay Second Derivative; MSC; Mean Centered; Average of triplicate measurements from each bioreactor).

Similar to the batch data, the MIR fed batch data do not display any signals of the distinct target peaks, glucose or lactate (Figure 4.25). Again this indicates that PLS modelling of the data is potentially not possible.

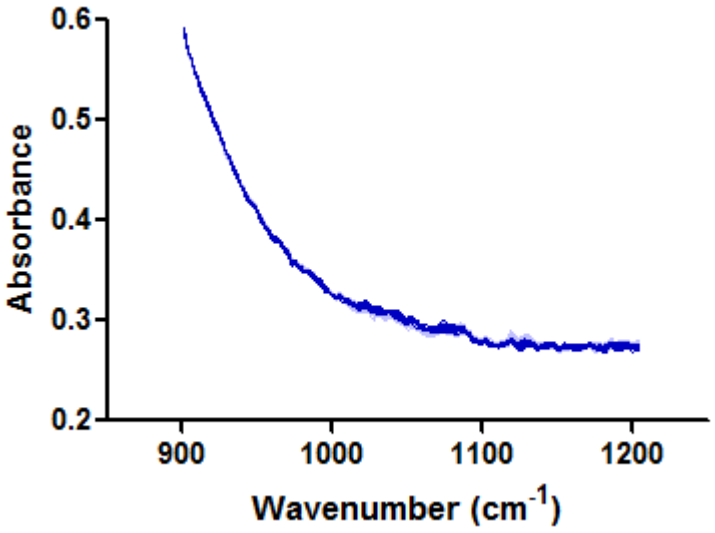


Figure 4.25: Preprocessed fed batch CD CHO MIR absorbance of 900-1200 cm^{-1} region from all processes [Colour Code Light to Dark = Bioreactor 1 & 2] (Processing: MSC; Mean Centered; Average of triplicate measurements from each bioreactor).

The Raman data are encouraging, with the two target peaks being more prominent once the appropriate pre-processing techniques were applied (Figure 4.26). Like the NIR fed batch data there is some difference between the batches, but these changes could represent the observations within the PCA plots.

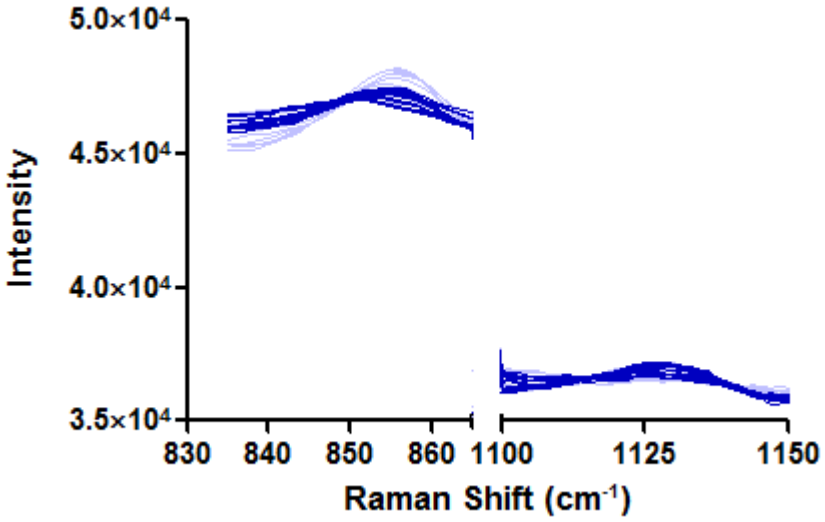


Figure 4.26: Processed fed batch CD CHO Raman scattering of 835-865 cm^{-1} and 1100-1150 cm^{-1} regions from all processes [Colour Code Light to Dark = Bioreactor 1 & 2] (Processing: MSC; Mean Centered; Average of triplicate measurements from each bioreactor).

4.2.3.6 Partial Least Squares (Calibration & Validation)

As with the batch data, the same PLS modelling procedure was performed upon the selected fed batch calibration and validation datasets selected. The fed batch PLS indicated, good model fit and low errors of prediction for both glucose (Figure 4.27) and lactate (Figure 4.28), comparable to those in the batch process. Although there is some shift in both the glucose and lactate models, which could be the result of inappropriate selection of samples in the prediction dataset.

[A]

	Calibration	Validation	Prediction
RMSE (g/L)	0.12	0.54	0.56
R^2	0.99	0.94	0.96

[B] **[C]**

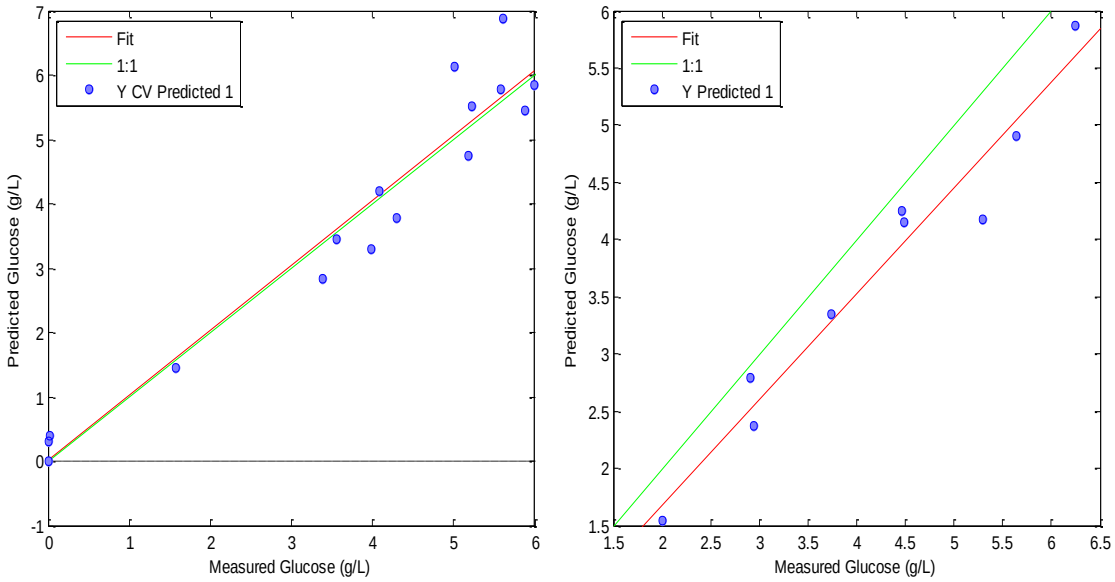


Figure 4.27: NIR Fed batch CD CHO glucose PLS [A] RMSE & R^2 values [B] Calibration & Internal Validation model [C] External Validation (Prediction) model; with the selected samples (blue points), the theoretical 'perfect' fit (green line) and the actual fit (red line).

[A]

	Calibration	Validation	Prediction
RMSE (g/L)	0.17	0.48	0.34
R^2	0.99	0.96	0.97

[B] **[C]**

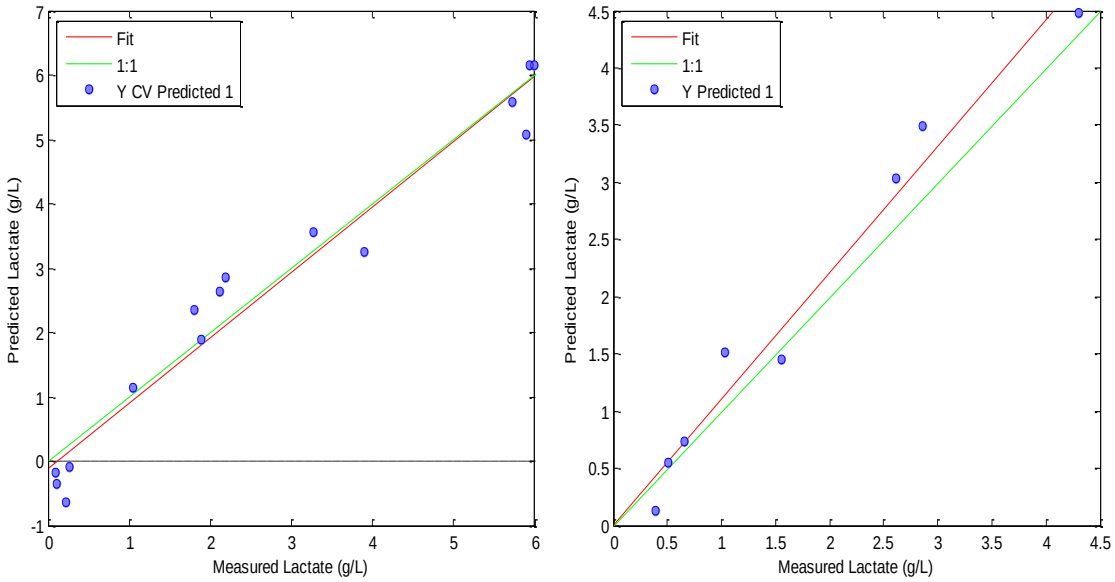


Figure 4.28: NIR Fed batch CD CHO lactate PLS [A] RMSE & R^2 values [B] Calibration & Internal Validation model [C] External Validation (Prediction) model; with the selected samples (blue points), the theoretical 'perfect' fit (green line) and the actual fit (red line).

The MIR PLS models and their RMSE for the fed batch processes reflect the results from the batch processes, with a poor model fit and high errors of prediction for both metabolites (Figure 4.29 & 4.30). The difference between the two batches, despite the trend between PC1 and PC2, has influenced the final model and even selection of calibration/validation samples to cover the plot space has not improved the model.

[A]

	Calibration	Validation	Prediction
RMSE (g/L)	0.02	2.07	1.45
R^2	0.99	0.07	0.49

[B] **[C]**

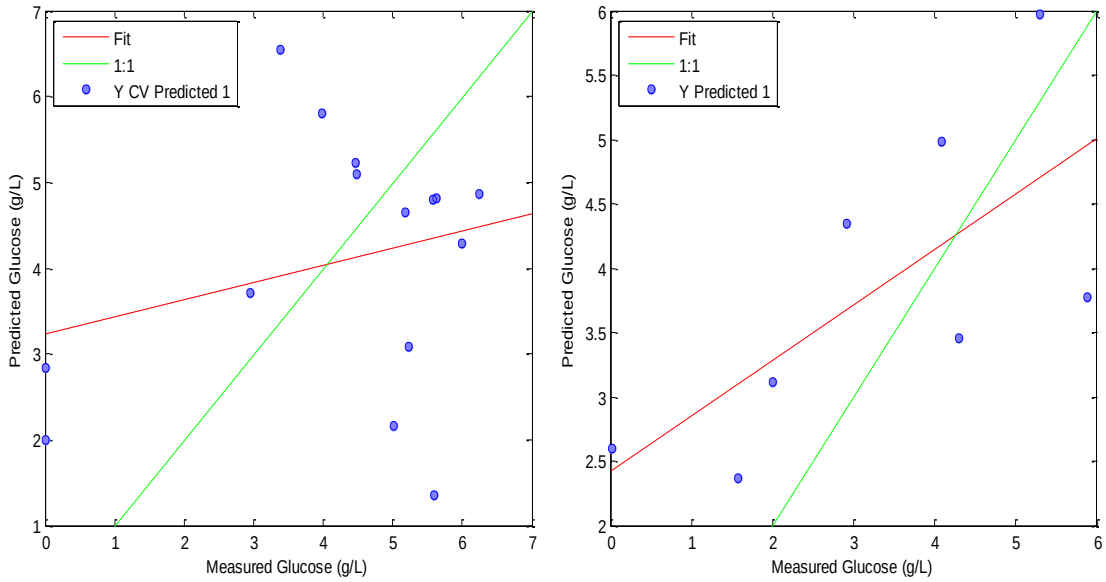


Figure 4.29: MIR Fed batch CD CHO glucose PLS [A] RMSE & R^2 values [B] Calibration & Internal Validation model [C] External Validation (Prediction) model; with the selected samples (blue points), the theoretical 'perfect' fit (green line) and the actual fit (red line).

[A]

	Calibration	Validation	Prediction
RMSE (g/L)	0.03	2.49	1.66
R^2	0.99	0.01	0.38

[B] **[C]**

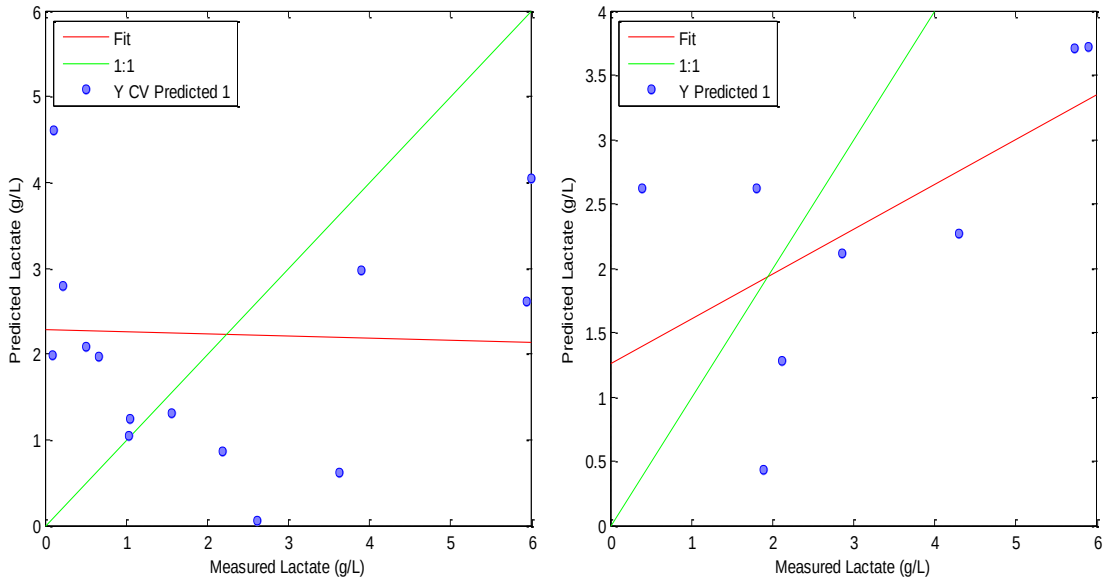


Figure 4.30: MIR Fed batch CD CHO lactate PLS [A] RMSE & R^2 values [B] Calibration & Internal Validation model [C] External Validation (Prediction) model; with the selected samples (blue points), the theoretical 'perfect' fit (green line) and the actual fit (red line).

Once again the Raman PLS models and their RMSE indicate a poor model fit and high errors of prediction for both glucose (Figure 4.31) and lactate (Figure 4.32).

[A]

	Calibration	Validation	Prediction
RMSE (g/L)	0.09	0.72	2.35
R^2	0.99	0.85	0.15

[B] **[C]**

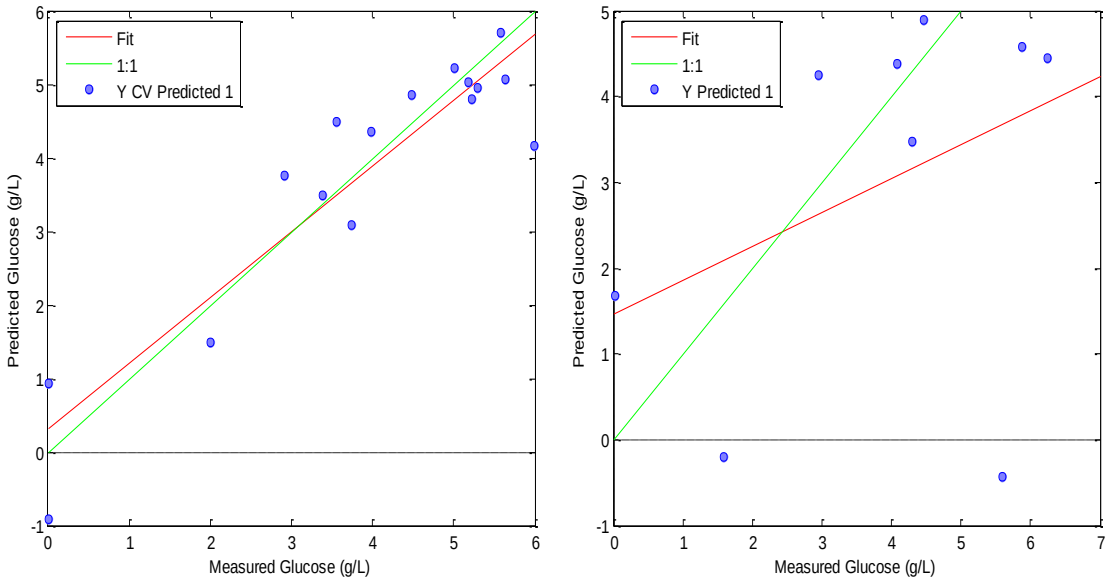


Figure 4.31: Raman Fed batch CD CHO glucose PLS [A] RMSE & R^2 values [B] Calibration & Internal Validation model [C] External Validation (Prediction) model; with the selected samples (blue points), the theoretical 'perfect' fit (green line) and the actual fit (red line).

[A]

	Calibration	Validation	Prediction
RMSE (g/L)	0.18	1.07	2.24
R^2	0.99	0.69	0.13

[B] **[C]**

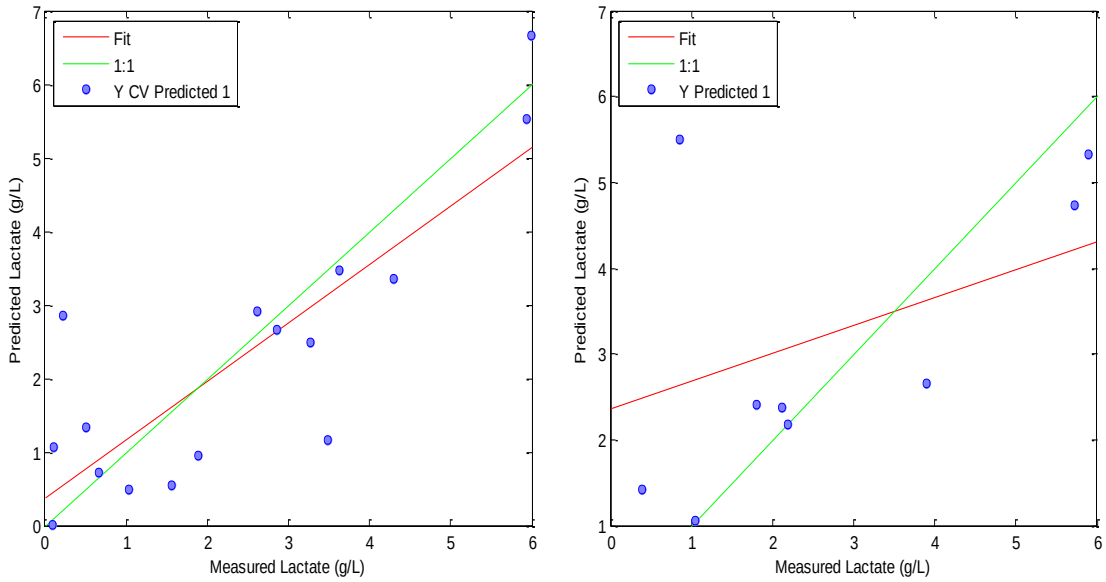


Figure 4.32: Raman Fed batch CD CHO lactate PLS [A] RMSE & R^2 values [B] Calibration & Internal Validation model [C] External Validation (Prediction) model; with the selected samples (blue points), the theoretical 'perfect' fit (green line) and the actual fit (red line).

Much like the batch results the PLS modelling of the fed batch processes has produced results of varying success. The models produced from the NIR once again indicate suitability for predicting both glucose and lactate. The MIR and Raman models however indicate poor suitability for predicting both metabolites.

4.3 Discussion

4.3.1 Spectral Models

4.3.1.1 NIR

As expected, based on the published literature, the prediction errors obtained using spectroscopic data in both processes were greater than those generated by the reference measurements (Hakemeyer *et al.*, 2012, Clavaud *et al.*, 2013). This is a result of the whole sample matrix influencing the prediction, hence the need to apply MVDA techniques to reduce the effects in the raw spectra (Biechele *et al.*, 2015, Rathore and Singh, 2015). These effects are absent in the YSI, as it utilises enzymes specific to only the metabolite of interest (3.9.3). However, despite the matrix complexity, it was possible to accurately predict both glucose and lactate in the batch (Figure 4.11 & 4.12) and fed batch (Figure 4.27 & 4.28) processes utilising the NIR models developed.

In terms of model predictive abilities both over- and under-fitting of models are undesirable, as they include not only relevant process data in the predictions but also noise, which is the result of including too many or too few latent variables (Rhiel *et al.*, 2002, Faber and Rajko, 2007, Gowen *et al.*, 2011). In regards to the batch data in this study, the glucose prediction for the external validation was superior to that of the internal validation, indicating that there was no over- or under-fitting of the calibration dataset. Unlike the glucose data, the lactate prediction is higher than the internal validation demonstrating an over fit in the initial calibration. For the fed batch processes the inverse of these results was observed, with the lactate prediction lower and glucose prediction higher, than the internal validation.

The results obtained for both processes are comparable and in some instances better to those from other reported CHO studies. Sandor *et al.* (2013) and Milligan *et al.* (2014) reported glucose predictions of 0.48 g/L and 0.99 g/L with R^2 values of 0.73 and 0.99, respectively. The values attained for within this study are comparable to those from the aforementioned authors, with the glucose predictions being 0.45 g/L (batch) and 0.56 g/L (fed batch), and the R^2 value of 0.96 (both processes). Likewise the lactate results from this study, RMSEP 0.34/0.45 g/L and R^2 0.97/0.94, in the batch and fed batch processes respectively also correlated well with the published data of Sandor *et al*, with an RMSEP of 0.44 g/L and R^2 value of 0.86.

The general increase in RMSEP value from the batch to fed batch processes could be the result of a combination of physical and metabolic changes in the different processes, as a number of studies have indicated that these can play have a significant influence upon the resulting models (Clavaud *et al.*, 2013, Sandor *et al.*, 2013, Milligan *et al.*, 2014). One factor that could be addressed to improve prediction is further investigation into the selection of the samples chosen for the calibration/validation datasets. However, the small size of the datasets in comparison to other published studies (Hakemeyer *et al.*, 2012, Clavaud *et al.*, 2013, Mercier *et al.*, 2016), limits the number of selection options in developing the calibration/validation datasets.

Overall the low errors of prediction in the batch and fed batch models for both metabolites clearly demonstrate the benefits NIR can offer to industry, as a process monitoring technique which provides predictions with an acceptable degree of accuracy.

4.3.1.2 MIR

Despite success in a previous study carried out by the author, utilising at-line MIR to analyse CHO cell culture samples, the results presented in this study do not indicate the

suitability of MIR as an analytical technique. One of the most plausible causes for this difference is the instrument. Previously an ATR zinc selenide (ZnSe) crystal based plate system was used, where as in this study an ATR Diamond Probe was employed. It has been highlighted that the composition of the crystal plays an integral role in MIR and that ZnSe is preferable to silicon or diamond based systems. This is because ZnSe offers a number of advantages such as, lower costs, higher refractive index, longer light paths and wider range of optical transmission transparency through the MIR spectrum (Roychoudhury *et al.*, 2006b, Sparks *et al.*, 2011, Reich, 2016).

The other potential cause of the MIR spectral results obtained in this study could be due to the poor temperature control, during the sample analysis. It has previously been reported that temperature variation has a significant effect on the spectrum, as the molecular bonds in the sample are affected by temperature variations (Foley *et al.*, 2012, Geörg *et al.*, 2015). As temperature variation was not accounted for during the sample analysis and the area where this took place was also poorly temperature controlled, it may have had a significant effect on the resulting spectra.

Although other studies have indicated the success of ATR MIR probes in bioprocess applications, including CHO cell processes, they do highlight some other issues associated with these systems that could have contributed to the results in this study. The limits of detection (LoD) of the instrument itself may have contributed to the lack of an apparent signal. It has been suggested that those concentrations of some metabolites in cell cultures are too low for certain instruments to analyse and the interactions within the complex media can lead to increased LoD (Foley *et al.*, 2012).

As previously discussed, the complex nature of media samples requires multivariate data analysis techniques to be applied to elucidate any information. However it has been

noted that as a result of this, a large sample set is advised and as the sample set in this study is only composed of two or three batches this could have had an influence in the poor model performance (Vojinovic *et al.*, 2006, Foley *et al.*, 2012).

The final issue that has been highlighted in some studies to impact on the spectra is the environmental conditions i.e. temperature. Control of these factors is crucial as they can lead to increased noise and signal drift in the spectra (Foley *et al.*, 2012, Geörg *et al.*, 2015). The results here demonstrate that extensive development and optimisation is required for MIR to be successfully applied to these CHO cell processes.

4.3.1.3 Raman

A number of recent studies have demonstrated the successful application of Raman spectroscopy for the analysis of CHO cell cultures and the subsequent quantitative modelling, incorporating the metabolites of interest in this study (glucose and lactate) (Whelan *et al.*, 2012, Berry *et al.*, 2015, Mehdizadeh *et al.*, 2015, Singh *et al.*, 2015, Matthews *et al.*, 2016).

Although the Raman spectra for both the batch and fed batch processes displayed visible activity in the spectral regions of interest indicated by the literature for glucose (1128 cm^{-1}) and lactate (855 cm^{-1}) (Mehdizadeh *et al.*, 2015, Singh *et al.*, 2015), the subsequent quantitative modelling of the data in this study proved challenging and ultimately unsuccessful.

A number of strategies were employed during the course of this study to not only try to improve the obtained spectra but also the resulting models. When the initial analysis was performed on the samples for both the batch and fed batch processes, there was significant variation observed between the spectra of sample replicates (data not shown).

Four parameters were investigated to decrease this variance and improve the raw spectral quality, laser power output, sample focal point and the exposure time.

The laser power output was the first parameter investigated as any fluctuations would indicate a serious fault with the instrument itself. A continuous measurement was carried out for an hour using a Spectra-Physics 407A laser power meter (Newport Corporation, CA, USA). The output of the laser power source was set at 400 mW, as per manufacturer's instruction, and the final readings from the laser probe head was consistently read as 200 mW. This 50 % loss from source to probe is acceptable as the loss of power is caused by the absorption and scattering of the light through the fibre, and consistent with the manufacturer's specifications. As there was no fluctuation in power output from the laser probe during these measurements, the instrument was ruled out as the cause of the spectral variation.

The next parameter investigated was the sample focal point because having the sample out of position would result in the spectra having increased noise and reduced intensity in the target peaks. The initial measurements were performed using the most concentrated sample from each process to find the focal point, as it was assumed that this would result in the optimal spectra and based on achieving a pixel fill between 50-70%. However as the instrument is calibrated/validated with cyclohexane, utilising this to optimise the focal point was investigated, as cyclohexane has a distinct Raman spectrum. It was discovered that by using cyclohexane, instead of the process sample to optimise the focal point, that the distance between the sample and the probe had to be adjusted and resulted in a reduction in variance between the replicates.

Once the focal point was optimised, the last parameter investigated to improve the spectra was the exposure time during the sample scanning and a range of exposure times

from 5 to 40 seconds were analysed. From these an exposure time of 35 seconds was selected as it increased the intensity of the target metabolite peaks, without increasing noise. Despite the significant improvements made on the Raman spectra these still yielded only qualitative data from the batch and fed batch processes.

The high errors of prediction for both metabolites, from both the batch and fed batch process indicate that further development and optimisation is required to utilise Raman spectroscopy in the presented CHO cell processes.

4.3.2 Data Fusion

As previously stated one of the aims of this study was to apply data fusion to the spectral data generated, with the goal of creating 'fused' models with lower errors of prediction than the individual parts. Unfortunately, this aim was not realised as only the NIR instrument produced data of a quantitative nature and therefore further investigation of data fusion had to be postponed. But the application of data fusion has had increasing interest over the past few years and does continue to offer a novel method of further developing spectral process models in the biopharmaceutical industry.

A number of studies have demonstrated the potential of data fusion to create a 'fingerprint' for the characterisation of various processes and products, such as craft beer and heavy oil (Biancolillo *et al.*, 2014, Laxalde *et al.*, 2014). This ability to create a unique 'fingerprint' for a process/product offers the biopharmaceutical the opportunity to not only characterise processes and detect deviations in product formation, but also detect counterfeit pharmaceuticals.

Although the application of data fusion in the biopharmaceutical industry has received very little attention and none in terms of biopharmaceutical process monitoring,

there has been some limited interest. One study applied a data fusion approach to four spectroscopic techniques for assessing raw materials used in mammalian cell culture. They demonstrated that while two techniques had a high predictive capability in terms of material/product quality, when all were 'fused' the single model produced greater predictions. The study offers a clear insight to the potential that data fusion can offer the biopharmaceutical industry, if taken beyond initial characterisation of raw materials and into process monitoring (Lee *et al.*, 2012b).

4.3.3 Batch vs Fed Batch

As all the research in this study was done in collaboration with Lonza and Thermo Fisher Scientific, meeting their needs into characterising the batch/fed batch systems across different media was essential. This particular study formed the baseline and first process development stage for the future studies, using Thermo Fisher Scientific's 'simplest' medium CD CHO.

The aim of a fed batch culture is to ultimately prolong process/production time to increase cell density and prevent cell death induced by nutrient depletion. However even with application of a feeding strategy, cell death can be induced by other factors such as; hyperosmolality, oxygen limitation, or accumulation of metabolic by-products i.e. lactate (Han *et al.*, 2011, Barrett *et al.*, 2012). With the aim of developing and improving the overall process a fed batch strategy was employed here, based up a protocol from a published study investigating a number of feeding regimes in different culture media (Barrett *et al.*, 2012).

The feeding strategy Barrett *et al.* (2012), outlined in the CD CHO medium used a similar proprietary IgG producing CHO cell line at a comparable scale. However, when their feeding strategy was applied to the processes in this study, it did not improve overall

process time, viable cell density or decrease lactate accumulation. In regards to the highest viable cell density the observed results, $5.8x10^6$/$5.2\ x10^6$ cells/mL, between the batch and fed batch cultures respectively, are not that different. In comparison to the study by Barrett *et al.* (2012) the highest viable cell density achieved here was roughly four fold lower than what they attained.

The initial difference in the glucose utilisation and lactate accumulation over the first seven days, between the batch and fed batch processes, could be due the reduced glucose conversation to lactate which has been reported in fed batch systems (Lee *et al.*, 2015). The decrease in lactate concentration in the latter stages of both processes indicated that the cultures may have switched metabolism. This switch has been reported to occur due to a number of reasons, such as reduced glucose or glutamine availability, which could be applied to the batch and fed batch culture respectively (Tsao *et al.*, 2005, Zagari *et al.*, 2013).

In terms of the batch culture the switch to lactate consumption could have occurred around day 8-9 as the lactate concentration shows a decrease and glucose is very limited. With the glucose concentration never reaching zero suggests that regeneration of glucose could be taking place through lactate consumption/conversation (Tsao *et al.*, 2005, Wilkens *et al.*, 2011, Zagari *et al.*, 2013). The notable increase in lactate on the final day of the process could also be a result of intracellular lactate being released from lysed cells, as it has been found that lactate can accumulate (Wilkens *et al.*, 2011). In the fed batch culture the addition of the feed may be responsible, firstly for the higher lactate concentration than that observed in the batch culture and the subsequent lactate consumption in the latter stages of the process (Zagari *et al.*, 2013).

Due to commercial sensitivity the specific concentration of hIgG produced cannot be disclosed, but the highest average concentration observed in the CD Batch process was within the concentration ranges reported in the published literature (Reinhart *et al.*, 2015a, Reinhart *et al.*, 2015b). However when transitioned into the fed batch process, the highest yield was half of that observed in the batch process.

While in theory this feeding strategy should have improved the initial batch process, the failure of the strategy could be down to the inhibitory effect of excessive feed which has been reported in CHO cell cultures (Han *et al.*, 2011, Lee *et al.*, 2015). Clearly, the transfer into a fed-batch protocol requires considerable optimisation of the process. Further bioreactor fed-batch runs would have been very useful in that process.

4.4 Conclusion

From the results presented in this study, it is apparent that only the NIR produced data of a sufficient quality to monitor metabolite formation in mammalian cell culture in both batch and fed batch processes, with the MIR being deemed unsuitable and the raman being suitable only in a qualitative approach. Thus investigating the feasibility of data fusion was not possible. The implementation of a fed batch protocol to improve the results of the initial batch process did not lead to any significant improvements and in some aspects adversely affected the process.

By utilising only the combination band region of the near infrared spectra successful construction of both glucose and lactate process models in both of the aforementioned process formats was achieved. These models produced errors of calibration, internal validation and external validation (prediction) which were comparable to those found in the literature.

Mid infrared models for both processes constructed produced very high errors across the calibration, validation and prediction. The poor performance of the models however reflects the results of the raw and processed spectra, where no apparent signals for either glucose or lactate were observed.

The Raman data demonstrated clear qualitative signals for glucose/lactate, in the target regions of the spectra where it was reported glucose and lactate are active. Unfortunately this did not translate into success when modelling these regions for the target metabolites, producing high errors, possibly caused by influences of other culture components in these regions.

Overall from the results achieved in this study NIR, for the purpose of glucose and lactate modelling, would be carried forth in the further development of the mammalian cell cultures in the more developed media. Despite only the qualitative data collected from the Raman, it would also be utilised in the next development stages.

Chapter 5:

Application of Dual At-Line Spectroscopic Monitoring to CD OptiCHO Batch and Fed Batch Chinese Hamster Ovary Cell Culture Systems

5.1 Overview

The next bioprocess development stage investigated in this study involved low passage number, batch and fed batch culture of the industrial CHO 42 mAb (hIgG) producing cell line, in the commercial chemically defined CD OptiCHO medium. Based upon observations in the published literature, these processes would provide suitable next developmental stages, in terms of spectroscopic characterisation of a CHO culture in a more developed medium (Barrett *et al.*, 2012, Reinhart *et al.*, 2015a). As with the previous study (Chapter 4) the low passage number was set at seven passages, as the industrial collaborators wished to investigate the requirements of early process medium in cultures under ten passages.

The strict regulatory stance within the biopharmaceutical industry has led to many innovations in terms of medium development. Initial focus was in developing media free from both animal components and serum for a number of reasons such as, variation in composition, potentially introducing microbial contaminants and the ethics of using these in human therapeutic production (Rodrigues *et al.*, 2013). With the success of serum-free media, focus has shifted into further development of the chemically defined media and feeds to improve cell culture processes (Yang, 2016), which is in part the purpose of this current study.

The conditions and sampling procedure outlined in 3.3 was utilised for both the batch and fed batch processes, and their replicates, performed during this study. *In-situ* measurements of pH, dissolved oxygen and temperature, along with at-line cell density and viability analysis, were performed to monitor the progress of each process. Metabolite and product analysis was split into either, reference (offline) or modelling (at-line). Offline measurements were analysed using the techniques and procedures outlined in 3.9, while at-line data was collected from the NIR and Raman spectrometers. To produce the final

metabolite models analysis of the spectral data was carried out using those the techniques outlined in 2.3/2.4.

5.1.1 Aim & Objectives

The aim of this study was to characterise the CHO cell culture, using a low passage number cell culture (seven), in the commercial CD OptiCHO medium in the batch/fed batch systems using a dual spectroscopic strategy of at-line NIR and Raman spectroscopy. These techniques were used to monitor glucose and lactate, with the objective of developing process models which could be compared to the CD CHO medium processes. By comparing the low passage cultures in both media the results will aid the industrial collaborators (Thermo Fisher Scientific & Lonza) in further developing an early stage process medium.

As these processes would be compared to the previous CD CHO processes, it was essential to fully characterise the CD OptiCHO processes in as similar a manner to the aforementioned processes as was possible. The results from the previous study demonstrated that it is possible to apply NIR and Raman spectroscopy for analysis of glucose and lactate, in a CD CHO medium culture and there have been reports of utilising these techniques in CD OptiCHO processes (Lee *et al.*, 2012a, Matthews *et al.*, 2016). These authors discuss the approaches they took to apply the respective spectroscopic techniques to CHO culture bioprocesses and the multivariate analysis strategies utilised to produce the metabolite models.

Similar to Chapter 4, both analysis of pure media samples and observations from the literature were used to determine the appropriate spectral regions to focus on, during subsequent sample spectral processing (Williams and Fleming, 1995, Colthup, 2012, Nyquist and Kagel, 2012, Smith and Dent, 2013). In the present study the previously outlined at-line NIR and Raman spectrometers would also be reassessed to determine if the instruments

had to be further optimised for data collection in the CD OptiCHO medium. From the collected spectral data and the off-line reference data, quantitative NIR models would be produced for glucose and lactate. Raman would also be used to produce qualitative information and reassessed for the development of quantitative models, for the two metabolites. The resulting models would be tested for suitability by subjecting them to both an internal and external validation, using selected samples from among the batches.

Overall this research represents the next stage of process development as it utilises a more complex medium. The results can be compared to the previous study to provide an insight into the performance and needs of a low passage number culture performance.

5.1.2 Novelty

As with the other CHO bioprocess investigated, these systems have not previously been investigated with a focus on characterisation of low passage number cultures. The process being investigated in this study also utilises a more complex chemically defined medium, which offers greater understanding in terms of what developments can be made in early culture media. This research provides further complexity and novelty within the investigated CHO bioprocesses, which can be fed back to the industrial collaborators.

Interpretation of spectral data has been well reported to be problematic due to spectral variance, influenced by a number of different factors. The CD OptiCHO medium in this study represents a significant change to the sample matrix, which could potentially render the previously identified data pre-processing approach from the CD CHO medium processes ineffectual. To determine if the CD OptiCHO medium led to any significant changes in the optimal data pre-processing, the novel DoE strategy previously employed was utilised again in combination with PCA in this study.

5.2 Results

5.2.1 Spectral Region Assignment

To determine if the change in medium composition, affects the position of the glucose and lactate signals in the NIR and Raman spectra, pure media samples were analysed. These results indicated that there was no effect on the metabolite signals and combined with the literature, the previously assigned regions for glucose and lactate were assigned to the NIR/Raman spectra (Table 5.1) (Chen *et al.*, 2004, Goodarzi *et al.*, 2015, Mehdizadeh *et al.*, 2015, Singh *et al.*, 2015).

Table 5.1: Specific peaks of target metabolites in CD OptiCHO medium.

Metabolite	NIR Peak Region	Raman Peak Region
Glucose	2200-2400 nm	1072 cm^{-1} and 1128 cm^{-1}
Lactate		855 cm^{-1}

5.2.2 CD OptiCHO Batch Process Analysis

5.2.2.1 Reference Analysis

Two, passage number seven CHO 42 batch processes in CD OptiCHO chemically defined medium were investigated to form a baseline for this chapter and as new development stage, through characterisation of the cell metabolism (glucose utilisation/lactate production), live cell density/viability (Figure 5.1) and hIgG concentration. The highest observed average cell concentration was on day 6 ($2.5x10^6$ cells/mL), with observed variations in viability the previous days and rapidly increasing lactate concentration. Subsequent glucose depletion and lactate accumulation resulted in the cultures being terminated on day 13, once the average live cell viability reached 6% (Figure 5.1). The hIgG

production was normalised as a percentage of the highest concentration achieved during the process due to commercial sensitivity, however it can be noted that the initial increase in concentration coincides with the rapid increase in cell density. Subsequent decrease in hIgG concentration during the final four days of the process is most likely to have been caused by degradation of the hIgG. See appendix Table A.3 for the corresponding values to Figure 5.1 and appendix Figure A.3 the bioreactor process condition values (pH, dO_2, Agitation & Temperature).

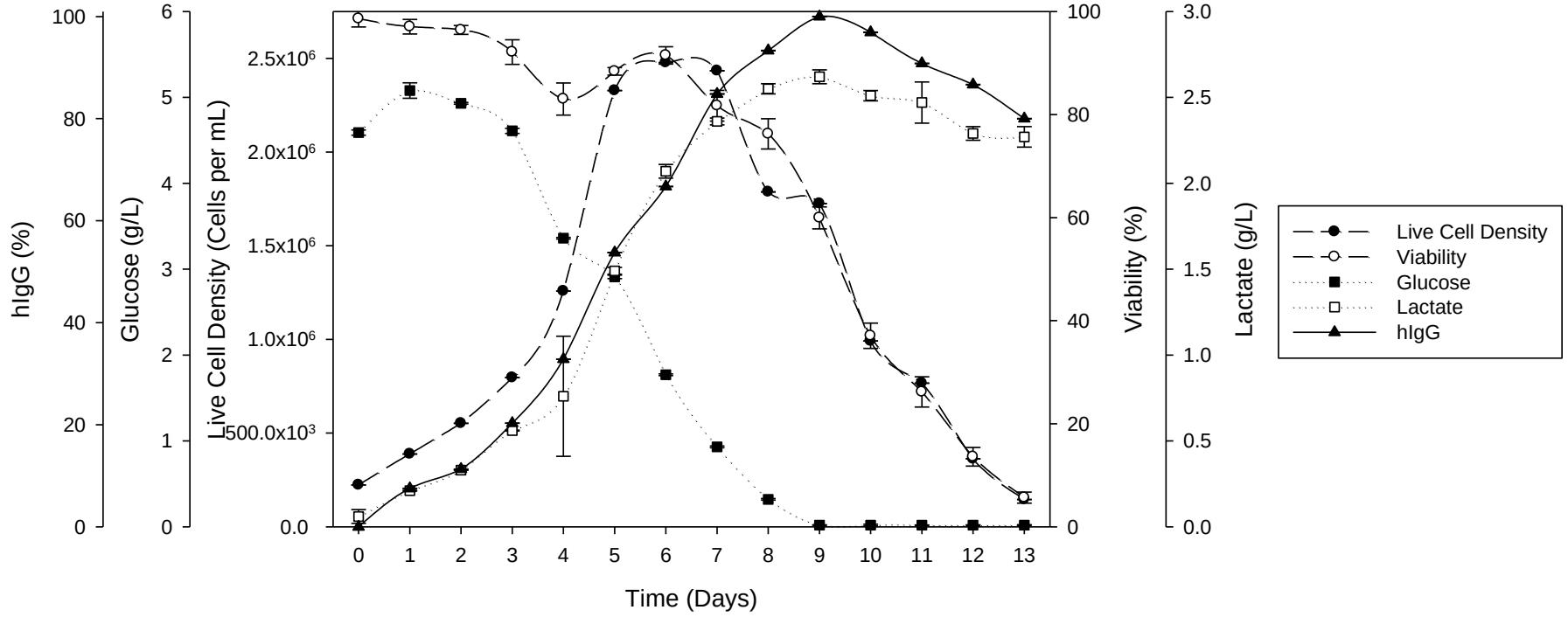


Figure 5.1: Time course of process data from CHO 42 culture in batch CD OptiCHO medium [average of triplicate measurements from two bioreactors plotted with standard deviation].

5.2.2.2 Raw Process Spectra

As reported in the literature, in the present study there were no visible signals present in the raw NIR process spectra (Figure 5.2), because of the dominant effect that the absorbance of water has upon the spectra (Beutel and Henkel, 2011, Reich, 2016).

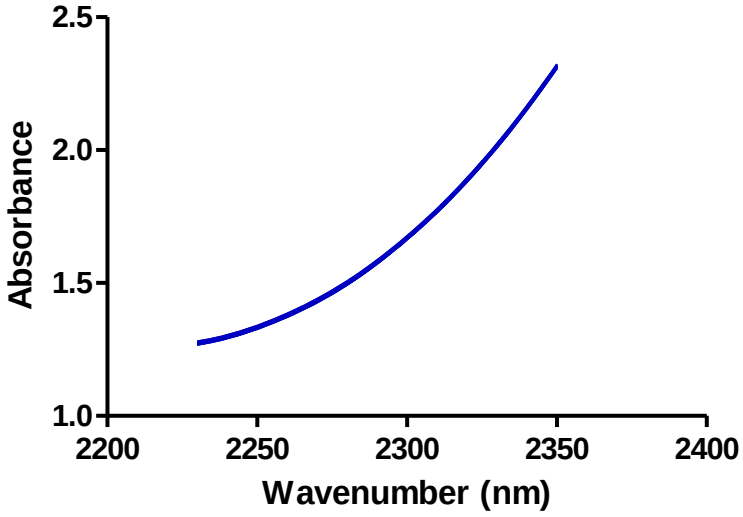


Figure 5.2: Batch CD OptiCHO NIR absorbance of combination band region (stretching and bending vibrations for C–H) from all processes [Colour Code Light to Dark = Bioreactor 1 to 2] (Processing: Average of triplicate measurements from each bioreactor).

From the Raman spectral data, the two peaks identified from the previous study and the literature, which corresponded to lactate (855 cm^{-1}) and glucose (1128 cm^{-1}), were identified in the process samples (Mehdizadeh *et al.*, 2015, Singh *et al.*, 2015). The regions around the two peaks, 835-865 cm^{-1} and 1100-1150 cm^{-1}, were utilised in the DoEman and PCA analysis to produce the process models (Figure 5.3).

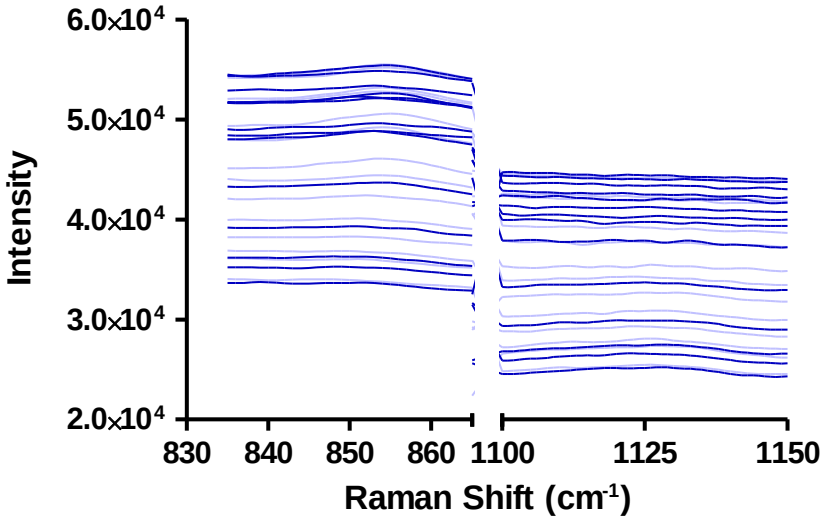


Figure 5.3: Batch CD OptiCHO Raman scattering of 835-865 cm^{-1} and 1100-1150 cm^{-1} regions from all processes [Colour Code Light to Dark = Bioreactor 1 to 2] (Processing: Average of triplicate measurements from each bioreactor).

5.2.2.3 DoEman

As previously highlighted, due to the change in medium, it could not be assumed that the optimised pre-processing treatments for the CD CHO medium were suitable for the CD OptiCHO medium. Using the DoE approach outlined in 3.8.2, a selection of pre-processing treatments were applied to determine what the optimal strategy for processing the data was (Table 3.2).

5.2.2.4 PCA Scores/Plots

Selecting the calibration and validation datasets was the next stage to optimise the spectral data and to do this PCA was performed upon the spectral regions of interest. The parameters outlined in Table 5.3, are those that resulted in the PCA capturing the maximum variance within both the NIR and Raman data.

Table 5.3: Spectral pre-processing PCA parameters and statistical values

Technique	Pre-processing	Cross Validation	No. of Latent Variables	Variance Captured
NIR	2nd Derivative, MSC & Mean Centering	Venetian Blinds	4	98.39 %
Raman	MSC & Mean Centering	Venetian Blinds	4	94.07 %

The parameters outlined in Table 5.3 for the NIR PCA, resulted in an observable trend within the PCA scores [PC1 vs PC2] (Figure 5.4). While both batches appear different, the early process samples in both batches correlated to PC1 and then shifted towards PC2 over time as the process continued. Based upon the observations, two thirds of the data which were equally distributed across the PCA plot were selected as the calibration dataset. The remaining third was then utilised in the external validation data set.

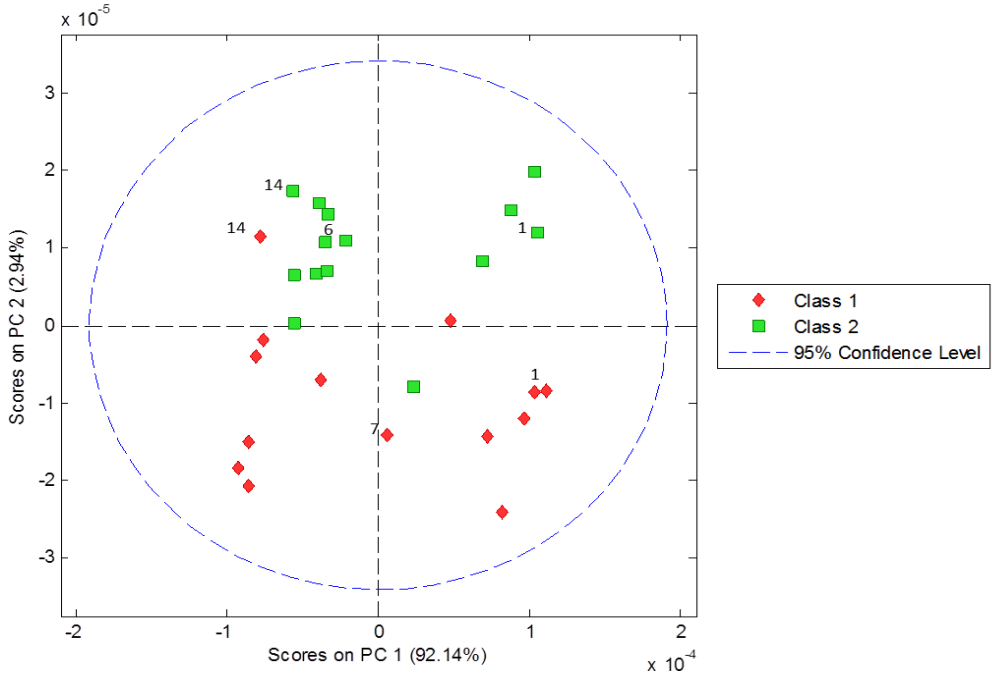


Figure 5.4: Batch CD OptiCHO NIR PCA scores of the combination band region in all processes, with process sample numbers represented [Class=Bioreactor 1-2] (Processing: Savitzky-Golay Second Derivative; MSC; Mean Centered; Average of triplicate measurements per batch).

The results of the Raman PCA (Figure 5.5) indicated a similar trend to that observed in the NIR, but unlike NIR there is more overlap indicating similarity in those samples. The earlier process samples again correlate to PC1 and then develop a correlation with PC2, as the process continues. As a result of the PCA, the calibration dataset was composed again of two thirds of the data and the other third of the data formed the validation dataset.

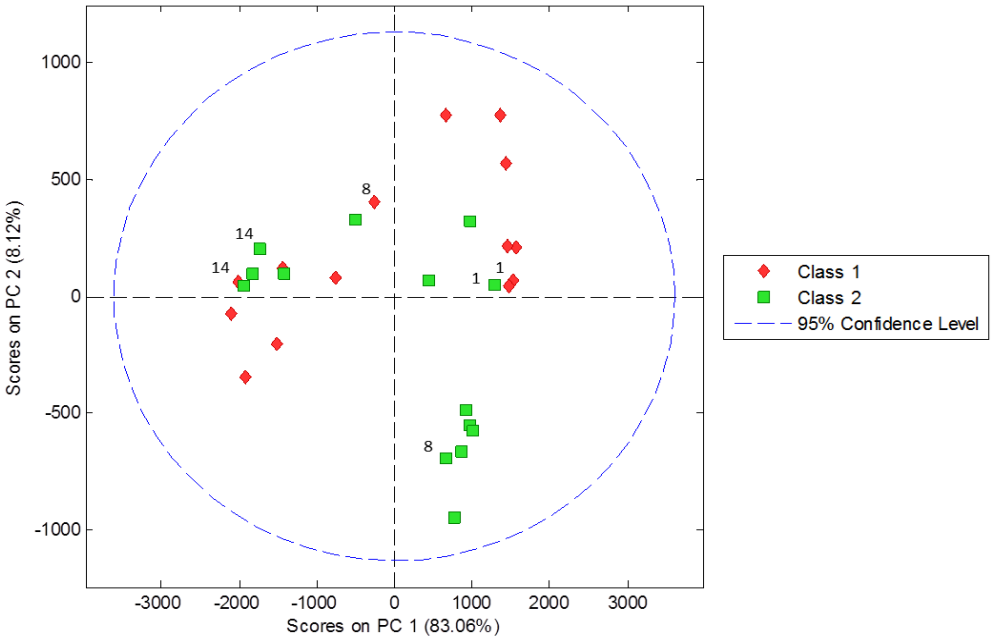


Figure 5.5: Batch CD OptiCHO Raman PCA scores of the 835-865 cm-1 and 1100-1150 cm-1 regions from all processes, with process sample numbers represented [Class=Bioreactor 1-2] (Processing: MSC and Mean Centered; Average of triplicate measurements per batch).

While it would be expected that the replicate batches presented should overlap in the PCA scores plots due to the similarity of the samples, this was not entirely clear in the present batches. Despite this, the similar observed trends within the PCA scores plots of each technique infer that PC1 and PC2 correlate to glucose and lactate, respectively.

5.2.2.5 Processed Spectra

In terms of changes to the raw NIR spectra once the optimised pre-processing techniques were applied, clear distinct peaks and troughs which correspond to the absorbance intensity within the CH region are now observable (Figure 5.6). The changes in height and depth of the peaks follow the concentration of glucose/lactate trends within the reference data from all batches. It is also notable that the intensity of the NIR absorbance peaks

differs between batch 1 and 2. This difference in intensity could reflect the observed differences between the batches in the previous PCA plot.

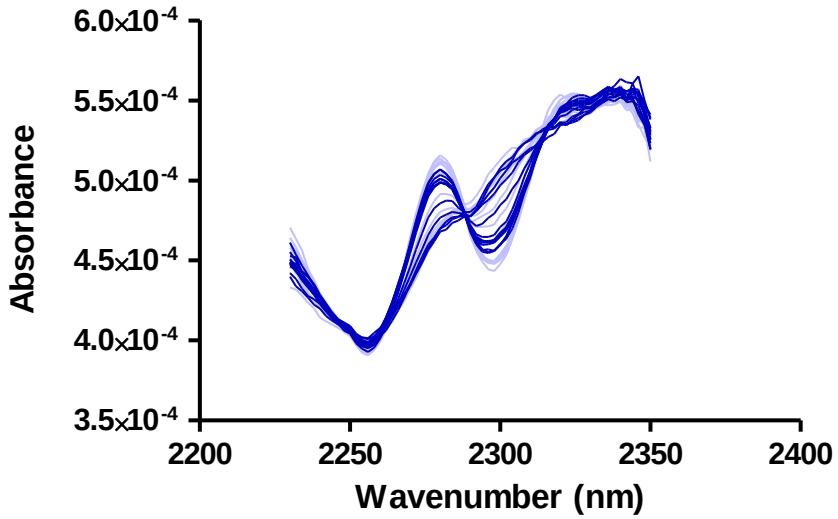


Figure 5.6: Preprocessed Batch CD OptiCHO NIR absorbance of CH region of all processes [Colour Code Light to Dark = Bioreactor 1 to 2] (Processing: Savitzky-Golay Second Derivative; MSC; Mean Centered; Average of triplicate measurements from each bioreactor).

The Raman spectral data after application of the optimised pre-processing treatments, like the NIR data, better elucidates the previously observed peak activity at the target spectral regions (Figure 5.7). The intensity of the peak at these regions correlates to the concentration of glucose and lactate within the samples of both batches.

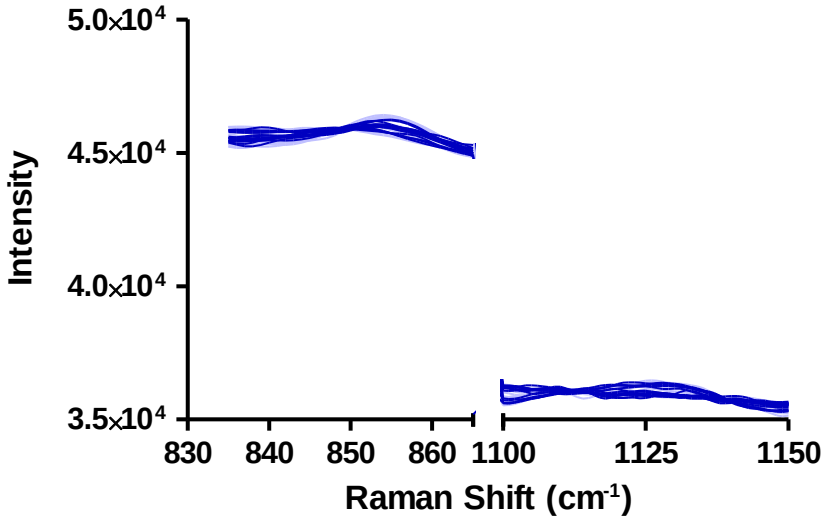


Figure 5.7: Processed Batch CD OptiCHO Raman scattering of 835-865 cm^{-1} and 1100-1150 cm^{-1} regions from all processes [Colour Code Light to Dark = Bioreactor 1 to 2] (Processing: MSC; Mean Centered; Average of triplicate measurements from each bioreactor).

5.2.2.6 Partial Least Squares (Calibration & Validation)

The final stage of the data analysis, after the spectral data optimisation was completed and calibration/validation datasets selected was to carry out the PLS modelling. The resulting PLS models and their Root Mean Square Errors (RMSE) produced from the NIR data indicated, a good model fit and with low errors of prediction for both glucose (Figure 5.8) and lactate (Figure 5.9). However there is an offset in the glucose model not present in the lactate model, which is possibly due to selection of samples in the prediction dataset.

[A]

	Calibration	Internal Validation	External Validation (Prediction)
RMSE (g/L)	0.17	0.33	0.37
R^2	0.99	0.97	0.98

[B] **[C]**

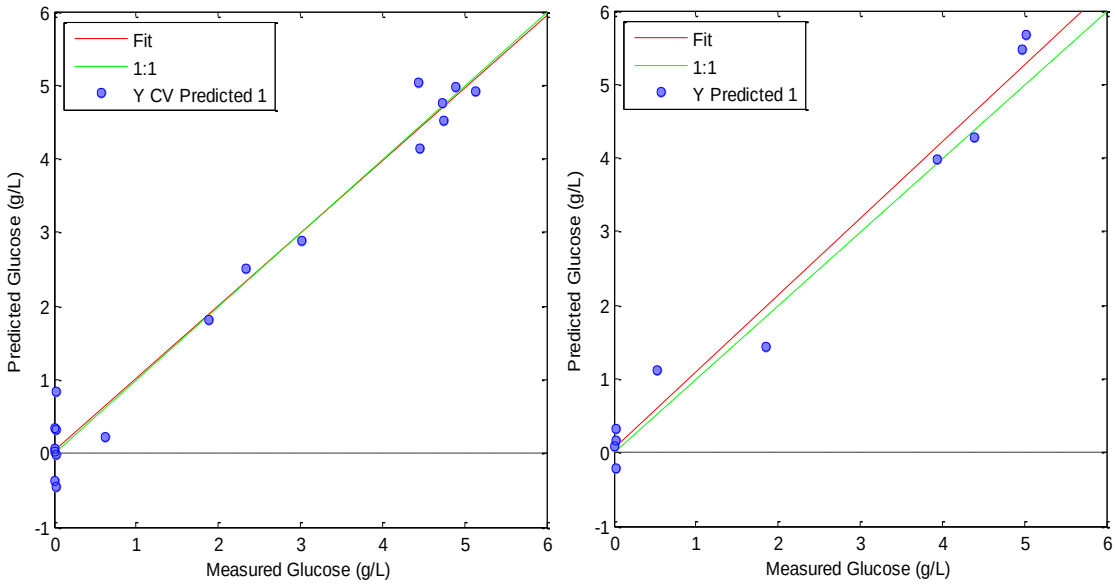


Figure 5.8: NIR Batch CD OptiCHO glucose PLS [A] RMSE & R^2 values [B] Calibration & Internal Validation model [C] External Validation (Prediction) model; with the selected samples (blue points), the theoretical 'perfect' fit (green line) and the actual fit (red line).

[A]

	Calibration	Internal Validation	External Validation (Prediction)
RMSE (g/L)	0.09	0.16	0.17
R^2	0.99	0.98	0.97

[B] **[C]**

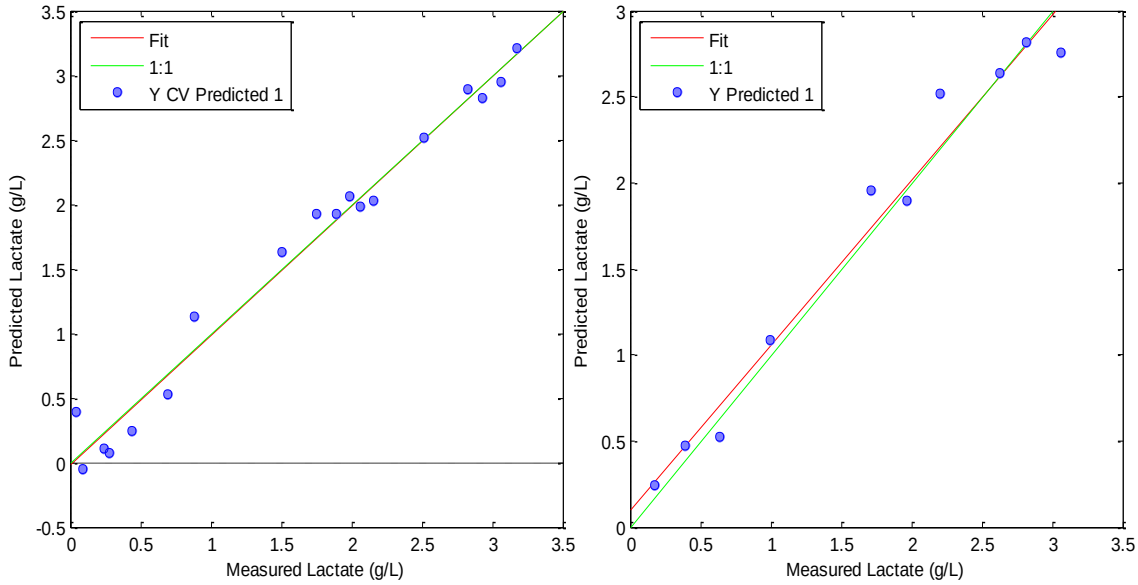


Figure 5.9: NIR Batch CD OptiCHO lactate PLS [A] RMSE & R^2 values [B] Calibration & Internal Validation model [C] External Validation (Prediction) model; with the selected samples (blue points), the theoretical 'perfect' fit (green line) and the actual fit (red line).

The Raman PLS models and their RMSE indicate a poor model fit and high errors of prediction for both glucose (Figure 5.10) and lactate (Figure 5.11). The model's poor predictive ability, despite the batches similar PCA plot trends, was to some extent anticipated (4.2.2.6) due to the previous study not demonstrating the success of quantitative Raman analysis. Undetected effects from other sample matrix components or poor selection of calibration/validation datasets, could potentially explain the poor model performance.

[A]

	Calibration	Internal Validation	External Validation (Prediction)
RMSE (g/L)	1.69	2.02	2.04
R^2	0.39	0.18	0.06

[B] **[C]**

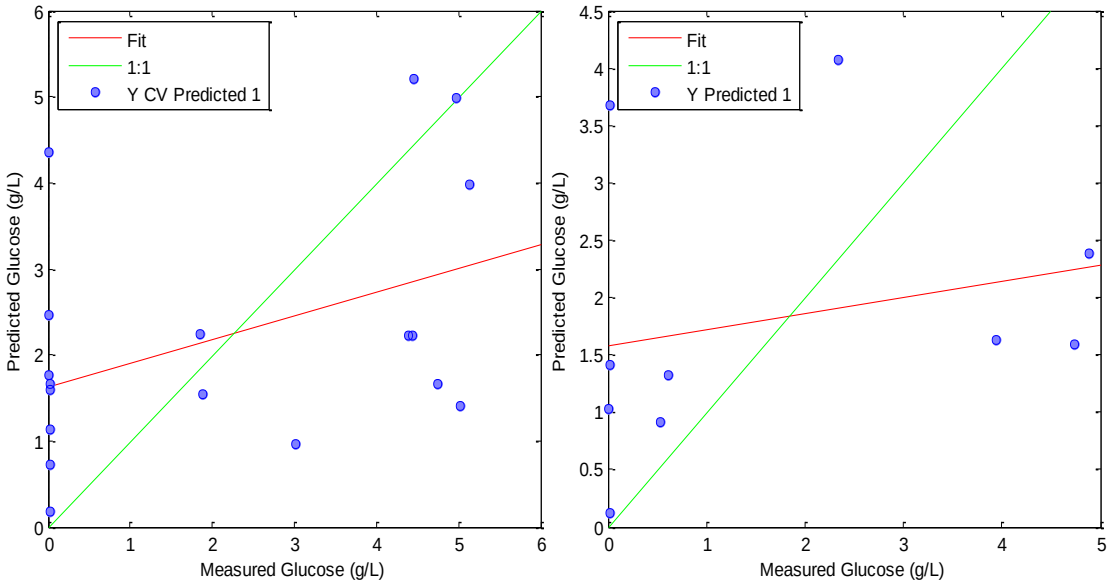


Figure 5.10: Raman Batch CD OptiCHO glucose PLS [A] RMSE & R^2 values [B] Calibration & Internal Validation model [C] External Validation (Prediction) model; with the selected samples (blue points), the theoretical 'perfect' fit (green line) and the actual fit (red line).

[A]

	Calibration	Internal Validation	External Validation (Prediction)
RMSE (g/L)	0.78	0.91	0.93
R^2	0.48	0.31	0.19

[B] **[C]**

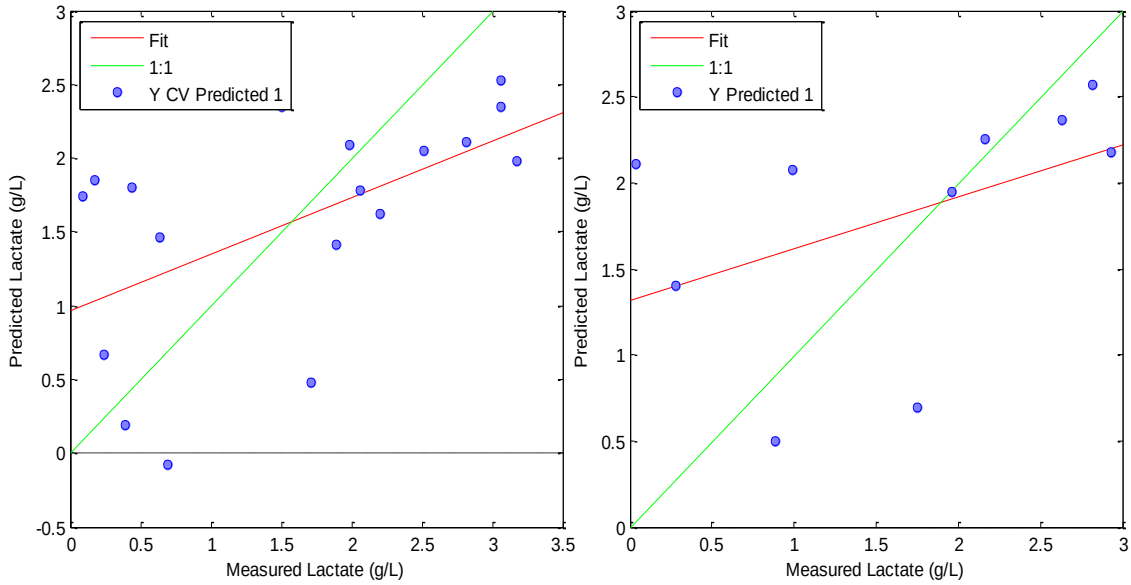


Figure 5.11: Raman Batch CD OptiCHO lactate PLS [A] RMSE & R^2 values [B] Calibration & Internal Validation model [C] External Validation (Prediction) model; with the selected samples (blue points), the theoretical 'perfect' fit (green line) and the actual fit (red line).

The PLS modelling has produced results which again indicate the suitability of NIR for predicting glucose and lactate. However, the Raman models indicate poor suitability for predicting both metabolites.

5.2.3 CD OptiCHO Fed Batch Process Analysis

5.2.3.1 Reference Analysis

To continue the process development, a fed batch process in the CD OptiCHO CHO medium with addition of CHO CD EfficientFeed™ A (EFA) was next to be carried out. The strategy

outlined in 3.2.2.1 was utilised and the processes characterised through the monitoring of the cell metabolism (glucose utilisation/lactate production), live cell density/viability (Figure 5.12) and hIgG production. Day 6 marked the highest cell concentration ($2.7x10^6$ cells/mL), slightly higher than that observed in the batch processes. However, this was subsequently followed by a sudden drop in viability, to 0% resulting in the termination of the culture. Despite employing a feeding strategy to improve culture life and performance, it has not been effected in a positive manner and the process was shortened significantly compared to the batch culture. Production of hIgG has been normalised as a percentage of the highest concentration achieved during the process due to commercial sensitivity. Once again the initial increase in concentration appears to be directly related to the increase in cell density, as does the subsequent decrease in the final day of the process which has led to potential degradation of the hIgG. See appendix Table A.4 for the corresponding values to Figure 5.12 and appendix Figure A.4 the bioreactor process condition values (pH, dO_2, Agitation & Temperature).

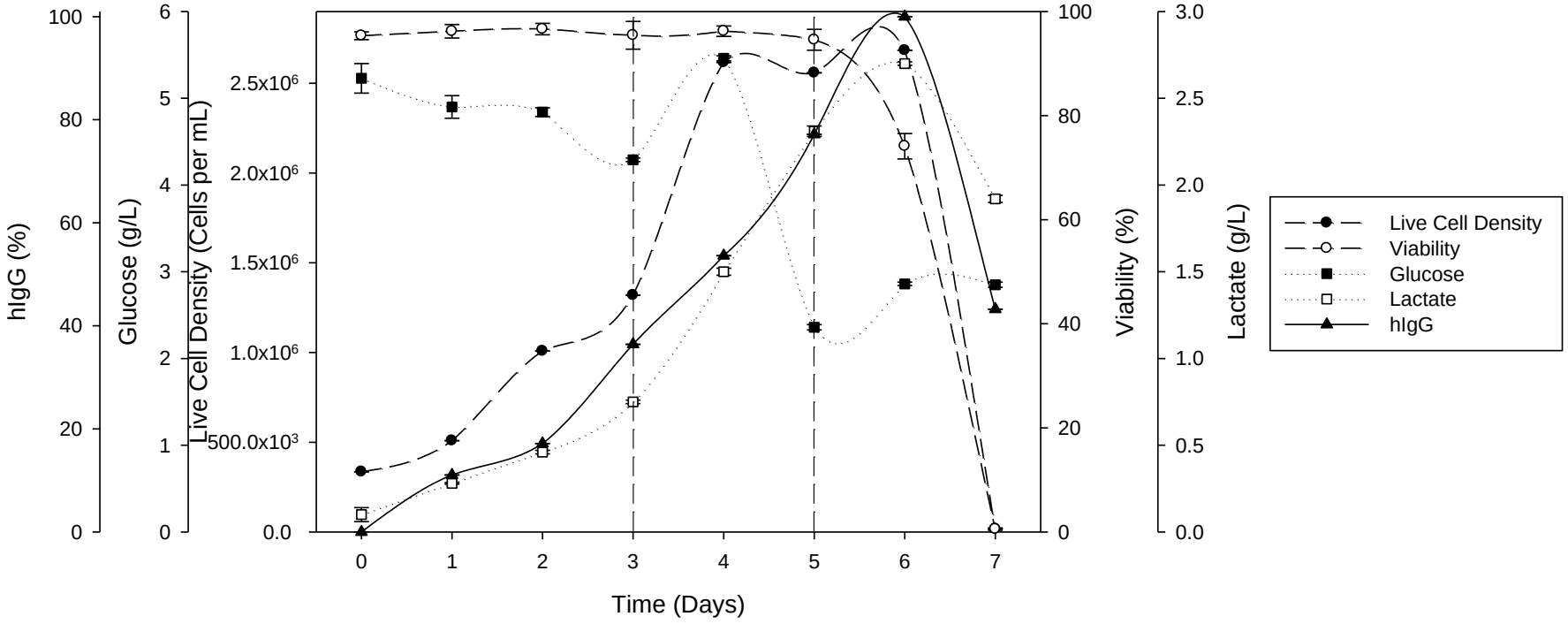


Figure 5.12: Time course of process data from CHO 42 culture in fed batch CD OptiCHO media. Vertical lines represent days when EfficientFeed™ A was fed to the culture [average of triplicate measurements from two bioreactors plotted with standard deviation error bars].

5.2.3.2 Raw Process Spectra

The fed batch NIR raw process spectra, like the batch processes, (Figure 5.13) did not exhibit any visible signals as a result of the intensity of the water absorbance.

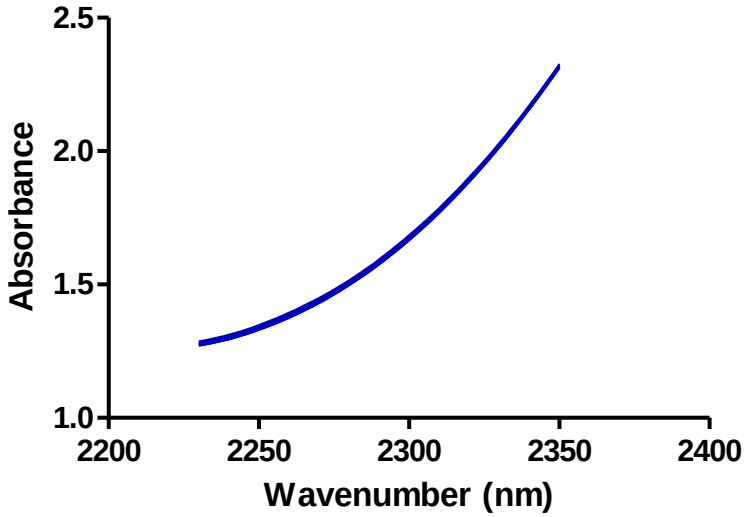


Figure 5.13: Fed batch CD OptiCHO NIR absorbance of combination band region (stretching and bending vibrations for C–H) from all processes [Colour Code Light to Dark = Bioreactor 1 & 2] (Processing: Average of triplicate measurements from each bioreactor)

The metabolite peaks for glucose (1128 cm^{-1}) and lactate (855 cm^{-1}) are once again present in the results from the Raman spectral data. But when compared to those peaks present in the batch cultures, they are much less pronounced. The regions around these two peaks, 835-865 cm^{-1} and 1100-1150 cm^{-1}, for lactate and glucose respectively were subsequently utilised in all future analysis (Figure 5.14).

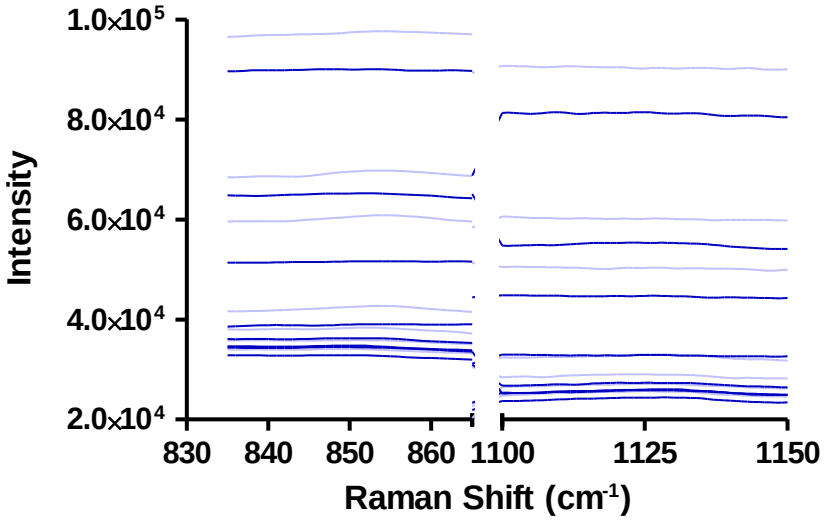


Figure 5.14: Fed batch CD OptiCHO Raman scattering of 835-865 cm^{-1} and 1100-1150 cm^{-1} regions from all processes [Colour Code Light to Dark = Bioreactor 1 to 2] (Processing: Average of triplicate measurements from each bioreactor).

5.2.3.3 DoEman

To improve and elucidate the features in the spectral data, the pre-processing optimisation strategy outlined in 5.2.2.3 was applied to the fed batch data for the future analysis.

The fed batch DoEman plots for both the NIR and Raman indicated that optimal pre-processing treatments were the same as those observed in batch processes analysis, for both glucose and lactate. The optimal preprocessing strategy for the NIR was indicated to be, 2nd Derivative (21, 2), MSC and Mean Centering. MSC and mean centering were identified as the optimised pre-processing treatments for the Raman dataset. The results of the DoEman were utilised in the PCA to define the calibration/validation datasets.

5.2.3.4 PCA Scores/Plots

Selection of the calibration and validation datasets was performed using PCA, to ensure spectral variation between the two batches was captured. Once the optimised pre-processing techniques were applied to the data in the PCA combined with an internal validation, the results which captured the maximum variance in the data were produced (Table 5.4).

Table 5.4: Spectral preprocessing PCA parameters and statistical values

Technique	Preprocessing	Cross Validation	No. of Latent Variables	Variance Captured
NIR	2^{nd} Derivative, MSC & Mean Centering	Venetian Blinds	5	99.15 %
Raman	MSC & Mean Centering	Venetian Blinds	5	96.15 %

From the results outlined in the NIR PCA (Table 5.4), there was a trend observed within the PCA scores [PC1 vs PC2] (Figure 5.15). In both batches the earlier process samples correlated to PC1 and as the process continues there was a shift towards a correlation to PC2. However both batches display some differences in the later samples, which corresponds to differences observed in the reference data. As a result of the PCA, two thirds of the data were selected equally distributed across both batches to form the calibration data set, with the remaining third forming the validation dataset.

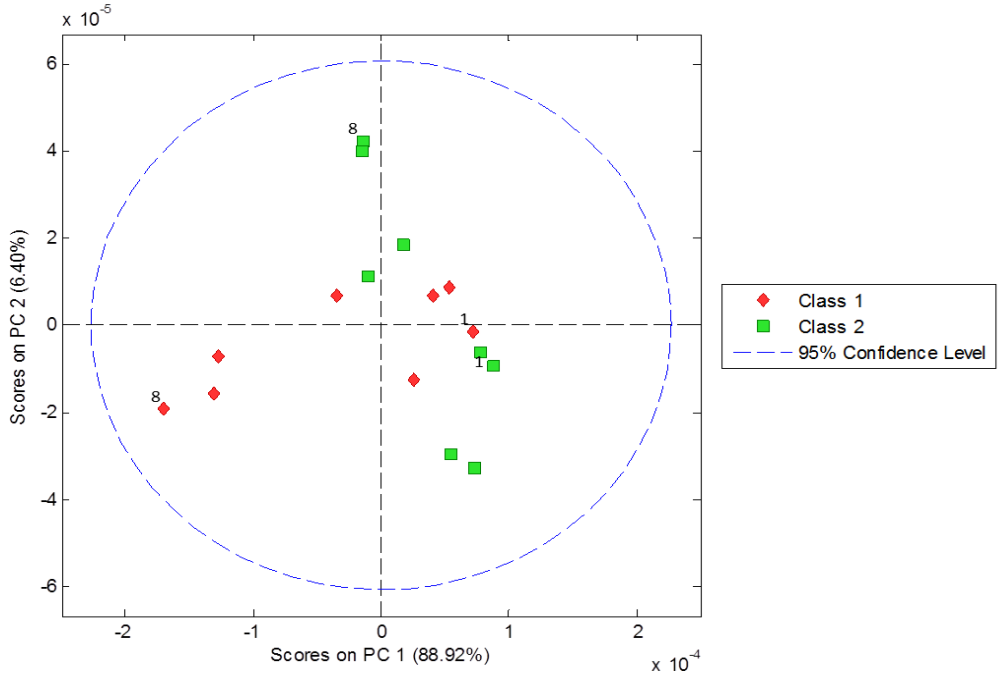


Figure 5.15: Fed batch CD OptiCHO NIR PCA scores of the combination band region in all processes, with process sample numbers represented [Class=Bioreactor 1-2] (Processing: Savitzky-Golay Second Derivative; MSC; Mean Centered; Average of triplicate measurements per batch).

The Raman PCA results (Figure 5.16) demonstrate the inverse trend of that observed in the previous datasets, the earlier process samples correlate to PC2 and then correlate to PC1 as process time increases. Once again there is only a minimal overlap between batches, at the beginning of the process and a notable difference between the batches in the latter samples. As a result of this, selection of the calibration/validation datasets was difficult in terms of selecting samples that cover the plot space. However again two thirds of the data was selected as the calibration dataset and the other third formed the validation dataset for the PLS modelling.

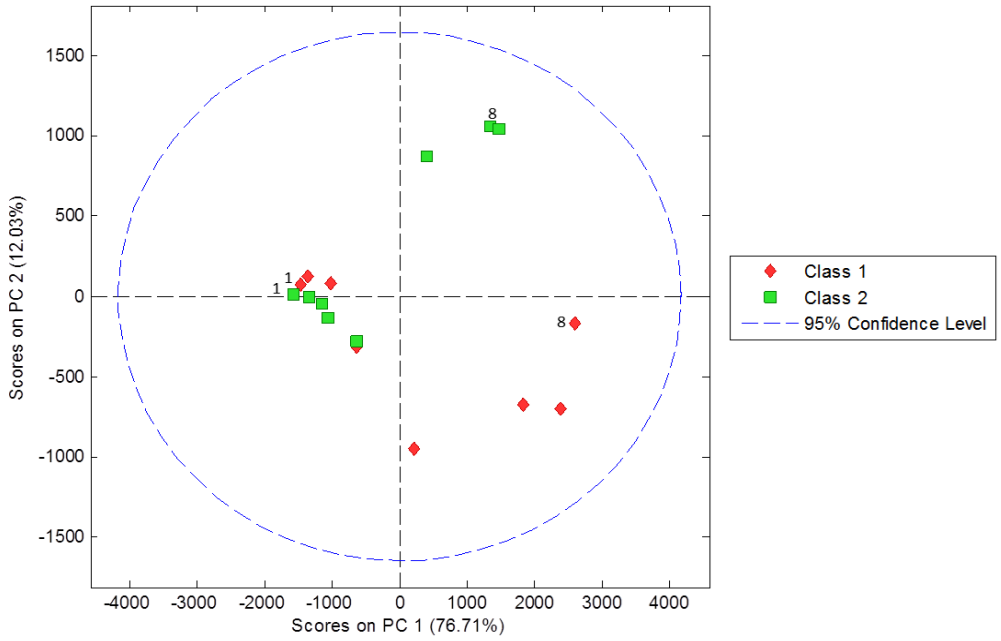


Figure 5.16: Fed batch CD OptiCHO Raman PCA scores of the 835-865 cm^{-1} and 1100-1150 cm^{-1} regions from all processes, with process sample numbers represented [Class=Bioreactor 1-2] (Processing: MSC and Mean Centered; Average of triplicate measurements per batch).

Despite there being minimal overlap in the fed batch PCA scores plots for both the NIR and Raman, indicating similarity between batches, the presence of the trends within the data indicate potential for modelling the metabolites. However the alternate trend in the Raman data, PC2 to PC1, indicates that modelling of this data would difficult in comparison to the NIR, due to the distribution of the samples.

5.2.3.5 Processed Spectra

As the optimised pre-processing treatments from the DoEman were applied to the fed batch spectral data distinctive peaks and troughs are observed, these reflect the glucose/lactate concentration trends within the reference data from all batches (Figure

5.17). But there is some difference in peak intensity between the two batches, which reflect the observed trends in the PCA plots.

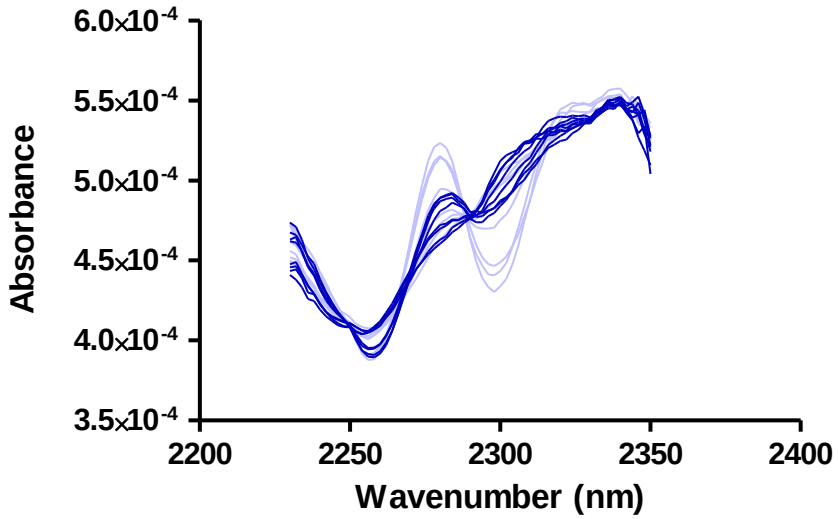


Figure 5.17: Preprocessed Fed batch CD OptiCHO NIR absorbance of CH region of all processes [Colour Code Light to Dark = Bioreactor 1 & 2] (Processing: Savitzky-Golay Second Derivative; MSC; Mean Centered; Average of triplicate measurements from each bioreactor).

The two target peaks are more distinct once the appropriate pre-processing techniques were applied to the Raman data (Figure 5.18). Similar to the NIR fed batch data, the previously observed difference in peak intensity between the batches can be seen more clearly than in the raw spectral data, these differences could explain the observations within the PCA plots.

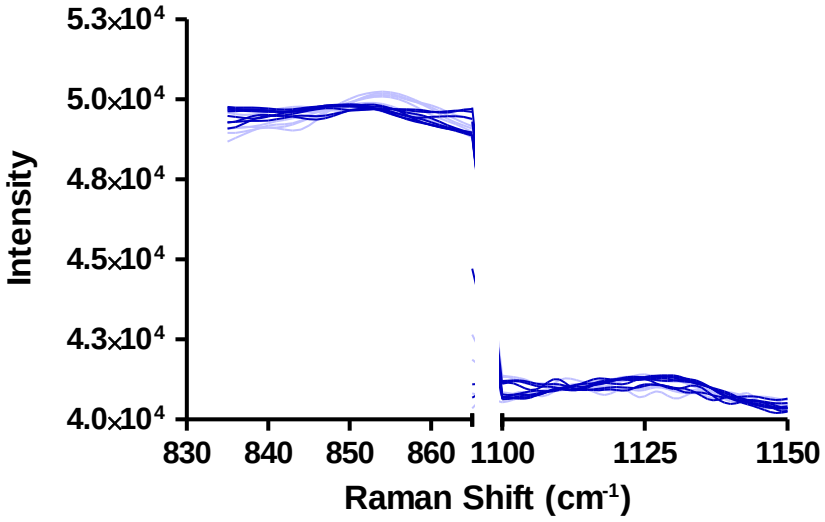


Figure 5.18: Processed Fed batch CD OptiCHO Raman scattering of 835-865 cm^{-1} and 1100-1150 cm^{-1} regions from all processes [Colour Code Light to Dark = Bioreactor 1 & 2] (Processing: MSC; Mean Centered; Average of triplicate measurements from each bioreactor).

5.2.3.6 Partial Least Squares (Calibration & Validation)

The same PLS modelling procedure was performed upon the selected fed batch calibration and validation datasets selected, as was done for the previous batch processes. In regards to the NIR PLS, the resulting models for both glucose (Figure 5.19) and lactate (Figure 5.20) demonstrate an acceptable fit and low errors of prediction, slightly higher than those observed in the batch process. The offset in both models, could however be caused by the samples selected in the modelling datasets.

[A]

	Calibration	Validation	Prediction
RMSE (g/L)	0.34	0.54	0.57
R^2	0.97	0.92	0.98

[B] **[C]**

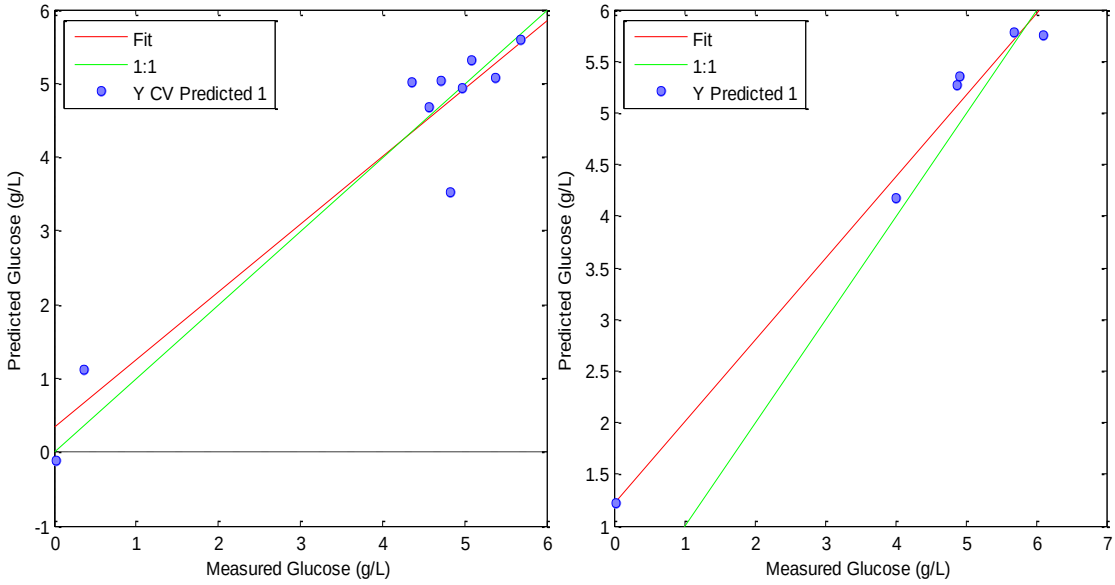


Figure 5.19: NIR Fed batch CD OptiCHO glucose PLS [A] RMSE & R^2 values [B] Calibration & Internal Validation model [C] External Validation (Prediction) model; with the selected samples (blue points), the theoretical 'perfect' fit (green line) and the actual fit (red line).

[A]

	Calibration	Validation	Prediction
RMSE (g/L)	0.07	0.11	0.33
R^2	0.99	0.99	0.96

[B] **[C]**

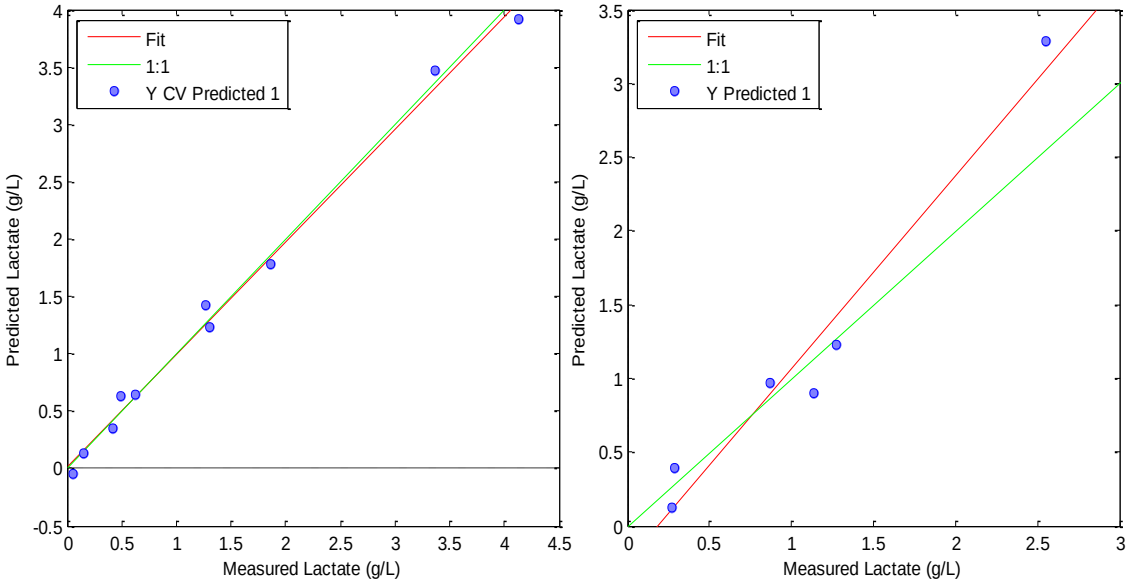


Figure 5.20: NIR Fed batch CD OptiCHO lactate PLS [A] RMSE & R^2 values [B] Calibration & Internal Validation model [C] External Validation (Prediction) model; with the selected samples (blue points), the theoretical 'perfect' fit (green line) and the actual fit (red line).

The Raman PLS models reflect those achieved for the batch processes, and the RMSE values indicate a poor model fit for both glucose (Figure 5.21) and lactate (Figure 5.22) and high errors of prediction.

[A]

	Calibration	Validation	Prediction
RMSE (g/L)	0.24	1.83	0.83
R^2	0.99	0.32	0.89

[B] **[C]**

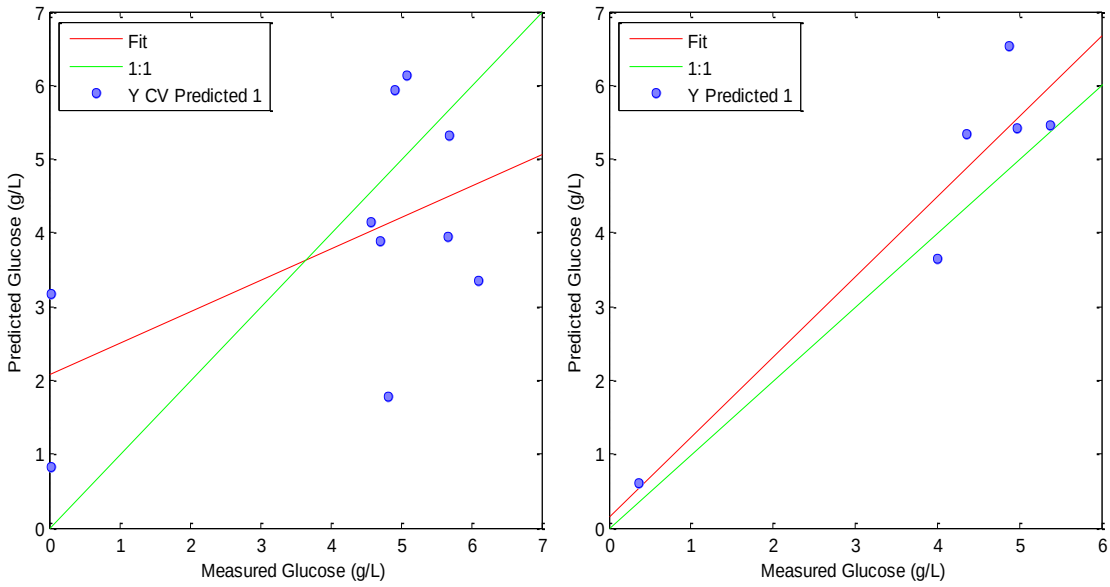


Figure 5.21: Raman Fed batch CD OptiCHO glucose PLS [A] RMSE & R^2 values [B] Calibration & Internal Validation model [C] External Validation (Prediction) model; with the selected samples (blue points), the theoretical 'perfect' fit (green line) and the actual fit (red line).

[A]

	Calibration	Validation	Prediction
RMSE (g/L)	0.08	0.74	0.24
R^2	0.99	0.64	0.96

[B] **[C]**

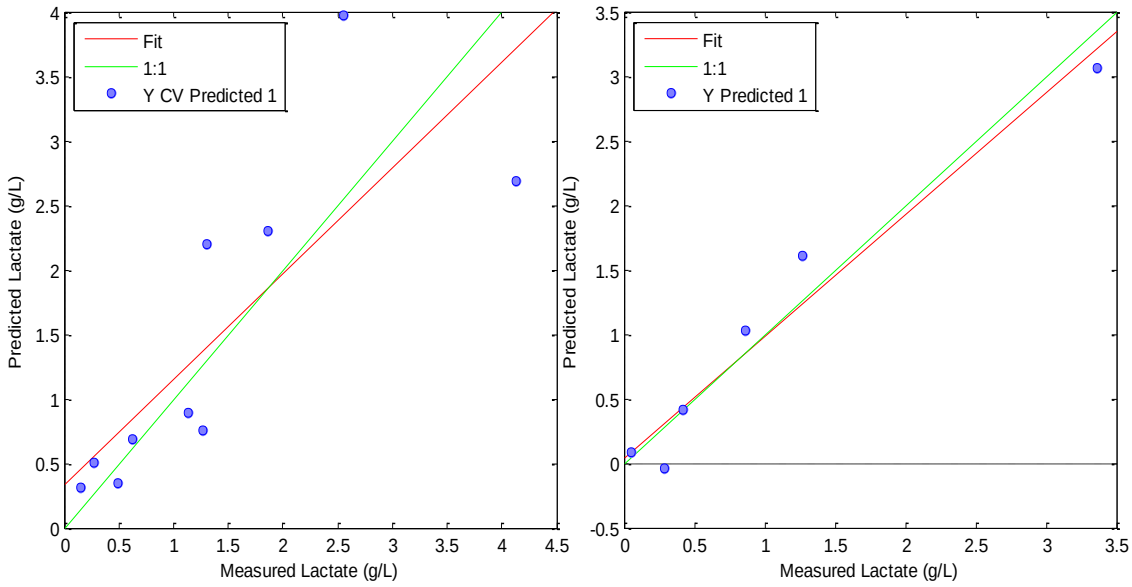


Figure 5.22: Raman Fed batch CD OptiCHO lactate PLS [A] RMSE & R^2 values [B] Calibration & Internal Validation model [C] External Validation (Prediction) model; with the selected samples (blue points), the theoretical 'perfect' fit (green line) and the actual fit (red line).

Overall the results of the fed batch process PLS models produced results of varying success. NIR produced models which displayed acceptable predictions for both glucose and lactate. Inversely the Raman models produced models which indicated poor suitability in predicting both metabolites.

5.3 Discussion

5.3.1 Spectral Models

5.3.1.1 NIR

As has been described previously, predictions using NIR may often produce greater errors than those from the reference methods, as a result of interactions in the sample matrix (Hakemeyer *et al.*, 2012, Clavaud *et al.*, 2013, Biechele *et al.*, 2015, Rathore and Singh, 2015). These effects are absent in the YSI, as it utilises enzymes specific to only the metabolite of interest (3.9.3). Even with the influences of these interactions, this study demonstrates the successful application of NIR in CD OptiCHO batch and fed batch cultures, to produce acceptable glucose (Figure 5.8 & 5.19) and lactate (Figure 5.9 & 5.20) prediction models.

For the batch data gathered in this study, the glucose and lactate errors of predictions for the external validation were greater than those of the internal validation, indicating that there was an over fit in the initial calibration. However the increases in error in the external validation were only minimal, 0.04 g/L and 0.01 g/L, for the glucose and lactate respectively. This observed over fit of the initial calibration was the same for the fed batch processes, with the glucose and lactate prediction errors being higher, 0.03 g/L and 0.22 g/L respectively, than the internal validation.

In terms of direct comparability to the published literature, there are only a very limited number of sources which have applied NIR to a CD OptiCHO based culture. However one study (Lee *et al.*, 2012a), assessed a number of selection algorithms in PLS modelling of a flask batch CHO cultures in CD OptiCHO medium. They reported a series of glucose prediction errors (RMSEP) ranging from, 1.15-1.45 g/L, using either 7 ±2 latent variables.

Even factoring in the slight overfitting of the prediction model, in comparison the glucose predictions produced in the present study using only 4 latent variables were far lower at, 0.37 g/L.

Due to limited application of NIR in OptiCHO processes, drawing direct comparisons between the results produced in the present study, for lactate and glucose/lactate in the batch and fed batch processes respectively and those in published works has not been possible. However, the results from these also indicate an acceptable level of error within the models. The lactate model produced from the batch culture data had an error of 0.17 g/L, which was lower than that observed in the glucose model. In regards to the fed batch models for glucose and lactate, they produced errors of 0.57 g/L and 0.33 g/L respectively, which were still lower than those observed in the study from the literature (Lee *et al.*, 2012a).

Based upon the NIR results from this study there is an acceptable degree of prediction errors for both metabolite models, which compare well to both the previous study (Chapter 4) and from the CHO culture literature. The low errors in prediction support the application of NIR to monitor an industrial bioprocess, as the technique could be partnered with a feeding strategy to counter the glucose depletion while also assessing lactate accumulation, which is detrimental to CHO processes.

5.3.1.2 Raman

Similar to the NIR, application of Raman spectroscopy to CD OptiCHO cultures in the published literature is limited. However, one very recent study investigated the application of a Kaiser Raman RXN2 spectrometer as a monitoring and control tool in a series of batch CD OptiCHO cultures at different scales (Matthews *et al.*, 2016). The models they generated through all the scales produced errors of 0.27 g/L and 0.20 g/L, for glucose and lactate

respectively. In comparison, the results from the batch processes in the present study were poorer, as they produced glucose errors of 2.04 g/L and lactate errors of 0.93 g/L, with very low R^2 values indicating poor correlation between the reference and spectral data. Although the systems used in both studies were the same, in the study from the literature (Matthews *et al.*, 2016) the system utilised different spectral acquisition strategy, which were not feasible in the present study. This strategy consisted of 600 accumulations of one second duration which took 17 minutes to acquire and the greater number of spectral accumulations used could have provided better defined spectra with reduced noise, making subsequent modelling more feasible.

Again there is limited application of Raman spectroscopy to similar CD OptiCHO processes in the literature, in particular for glucose/lactate modelling in fed batch processes, which has meant that gaining a comparison of the results in the present study has not been possible. Based upon the results produced in the study, it is safe to assume that the models produced here would be poorer than those in any published literature, indicated by the RMSEP and R^2 values.

While there are observable changes at the target peaks, 1128 cm^{-1} (glucose) and 855 cm^{-1} (lactate), in the Raman spectra from both batch and fed batch processes, even applying the developments outlined in 4.3.1.3 has failed to produce successful quantitative models for both metabolites. The results achieved here indicate that at this time the Raman modelling requires significant development to meet the needs of a PAT tool in these processes, as the high errors of predictions could not be used to monitor an industrial bioprocess.

5.3.2 Batch vs Fed Batch

This study represented a new stage in development, in terms of the needs of the industrial collaborators, Lonza and Thermo Fisher Scientific, by developing a batch/fed batch CHO culture system, using Thermo Fisher Scientific's more complex chemically defined CD OptiCHO medium. As previously described in 3.2.2.1, the feeding strategy applied to this study was from Reinhart *et al.* (2015) as they utilised a comparable IgG producing CHO cell line in CD OptiCHO medium.

The balance between over and under feeding is key in the fed batch cell process, as either can lead to a metabolic change. In terms of underfeeding, studies have shown that the glucose concentration should be maintained at 2-3 g/L and not be allowed to decrease below that (Barrett *et al.*, 2012, Lu *et al.*, 2013, Baik *et al.*, 2015, Reinhart *et al.*, 2015a). The issue with depletion of glucose or glutamine in CHO cultures is that without these components the cells have no precursors to feed into the TCA cycle for energy. As a result of this depletion a metabolic switch occurs where lactate is then utilised as an alternative precursor for the TCA cycle (Wilkens *et al.*, 2011, Zagari *et al.*, 2013). Counter to this, overfeeding poses just as serious an issue to the CHO cell culture as underfeeding. Overfeeding has been shown to affect the overall flux of glucose in the TCA cycle by upregulating the enzymes responsible for the conversation of pyruvate to lactate, to favour lactate production and the subsequent accumulation of lactate is in turn toxic to the culture (Lee *et al.*, 2015). It is possible that these reported metabolic observations could explain the difference between the batch and fed batch results presented in this study.

Although the fed batch strategy should have had a superior performance to that of the batch process, it has been clearly demonstrated that this was not the case. The highest viable cell density observed between the two process types was not altered by much,

$2.5x10^6$ cells/mL in the batch culture compared to, $2.7x10^6$ cells/mL in the fed batch culture. In contrast Reinhart *et al.* (2015) observed a peak viable cell density of $3.9x10^6$ cells/mL, a less complex flask fed batch CD OptiCHO process compared to the systems used in the present study. The most obvious difference in between the two process types is, of course the process time, as the fed batch process only ran for half the time observed in the batch process. The poor performance of the CD OptiCHO fed batch culture could be the result of a potentially inhibitory effects caused by excessive feeding (Han *et al.*, 2011, Lee *et al.*, 2015).

Another notable feature of both cultures is the decrease in lactate concentration in the latter stages of both processes which indicate a metabolic switch. However, this observation may not be applicable in the fed batch culture, as the change only occurred on the final day as the process was terminated. The lack of glucose and glutamine availability has previously been demonstrated to be influential in this metabolic switch and could potentially be the reason for this observed change in the batch and fed batch cultures respectively (Tsao *et al.*, 2005, Zagari *et al.*, 2013).

The final four days of the batch culture indicate a switch to lactate as a result of the glucose depletion, due to the observed decrease in lactate concentration. As the glucose concentration does not reach zero during the process, this again infers a switch to lactate utilisation with it being converted to glucose and then being rapidly utilised (Tsao *et al.*, 2005, Wilkens *et al.*, 2011, Zagari *et al.*, 2013). In terms of the fed batch culture, the switch to lactate at the end of the process may have been due to the addition of the feed, but due to this only being observed in the final day this may or may not be the case.

Overall the results obtained in this study indicate that transition of batch to fed batch culture process requires significant further development and optimisation. Only by performing more bioreactor fed-batch runs can these changes be achieved and any

meaningful conclusions regarding the performance of low passage number cultures in CD OptiCHO medium can be drawn.

5.3.3 CD CHO vs CD OptiCHO

In terms of comparing the results between the two medium processes from this study and the previous study, as the feed used between the two studies was different only the batch processes can be directly compared, while indirect observations can be drawn between the fed batch processes. Although the length of the batch cultures is the same in both media, the CD CHO medium produced a higher viable cell density of $5.8x10^6$ cells/mL compared to, $2.5x10^6$ cells/mL. This increase in viable cell density also resulted in the highest average hIgG yield observed, to be three times higher in the CD CHO batch culture compared to that achieved in the CD OptiCHO batch culture.

Since the Raman models for both the previous study and the current study were unsuccessful, only the NIR results will be discussed. The NIR model values from both studies are comparable, the glucose models produced RMSEP's of 0.45/0.37 g/L, and R^2 values of 0.96/0.98, for the CD CHO and CD OptiCHO respectively. The lactate results demonstrated an RMSEP value of 0.45 g/L for CD CHO and 0.17 g/L CD OptiCHO, with the models having an R^2 of 0.94 and 0.97 respectively. While the results are comparable, the RMSEP values from the CD OptiCHO cultures are twice as low as those observed in the CD CHO batches, indicating that the models perform better. This increased error in the CD CHO batches, could be due to there being three batches introducing more variance, whereas there were only two batches in the CD OptiCHO processes. As the feed for this fed batch process is different from the previous study (4.2.3), only an indirect comparison will be made between them. Looking at the NIR model values from both studies they appear to be very

similar. The models for both the glucose and lactate produced in the CD CHO and CD OptiCHO had RMSEP's which differed by only 0.1g/L.

As previously discussed, one theme of this and the previous study was to compare the performance of low passage number CHO cell cultures in the various media provided by Thermo Fisher, to better understand the medium needs in early process development. It has been well reported that high passage number has a number of negative effects upon CHO culture stability, affecting protein expression, cell growth and cell viability. These effects can be a result of loss of genes or mutations, leading to varying production rates between cell populations. The genetic change of some cell populations in the culture can potentially offer advantages which could therefore overgrow the initial production cell (Bailey *et al.*, 2012, Beckmann *et al.*, 2012, Veith *et al.*, 2016).

In contrast there is relatively little investigation into lower passage number cultures and their media requirements. Therefore only assumptions and speculation, based upon current understanding of animal cell physiology in culture systems can be made about how these cultures perform. Commercial cell lines go through multiple rounds of clonal selection and development to ensure stability, in terms of production, growth and medium suitability (Bailey *et al.*, 2012). It could be assumed from this observation that the transfer of a low passage cell culture designed for use in a medium containing a broad range of nutrients such as CD CHO, to a more specialised refined medium (CD OptiCHO) is not possibly and requires the initial development to be performed in the same medium.

From the results of the two studies, it appears that the CD CHO medium is better suited to early culture development due to the increased viable cell density and subsequent hIgG production. As both media are proprietary the differences between these is not fully

known, but it is clear that changes in medium composition have had a significant effect when utilised in low passage cultures.

5.4 Conclusion

The outcomes of the present study demonstrate the suitability of NIR as a qualitative and quantitative means of monitoring a mammalian cell culture in batch/fed batch processes, and the application of Raman as a qualitative technique for glucose and lactate analysis. The fed batch strategy used in this study proved to be a significantly detrimental development of the initial batch process. Rather than improving the results it had a detrimental effect upon the process by, increasing the decline in live cell density leading to a halving of the overall process time.

Application of near infrared spectroscopy in both batch and fed batch processes resulted in successful models for glucose and lactate, with low RMSE and R^2 values being produced. Despite the limited reporting of NIR in CD OptiCHO medium, there was some indication that the errors for glucose reported in some studies were more than double of those achieved in this present study. While the Raman data produced observable qualitative signals in the target regions corresponding to glucose and lactate, this did not lead to successful modelling of the data.

In conclusion the results of this study indicate that in comparison to the CD CHO medium, the CD OptiCHO medium is not as suitable for low passage cultures, as it produces lower viable cell densities and hIgG concentration. NIR is acceptable for glucose and lactate modelling and utilised in the final developmental stage of the mammalian cell cultures, in the Dynamis medium, along with the Raman for qualitative analysis.

Chapter 6:

Characterisation of Dynamis Batch Chinese Hamster Ovary Cell Culture by Application of Dual At-Line Spectroscopic Monitoring Strategy

6.1 Overview

The final bioprocess development stage to be investigated was a batch culture of the industrial CHO 42 mAb (hIgG4) producing cell line, at a low passage number, in the commercial Dynamis medium. This process was selected as the final developmental stage as the industrial collaborator, Thermo Fisher Scientific, had very little data regarding the performance of industrial cell lines in this medium.

The drive to develop chemically defined cell culture medium over the past two decades has resulted in several formulations, which have been shown to provide varying degrees of success (Barrett *et al.*, 2012, Reinhart *et al.*, 2015a, Ling *et al.*, 2015). Chemically defined culture media can have anywhere from fifty to one hundred different constituents, each of which plays a vital role. Even the more common medium constituents, such as glucose, can be further optimised in terms of concentration to improve process performance (Ling *et al.*, 2015). Dynamis medium represents Thermo Fisher Scientific's most recent effort at producing a complex chemically defined medium, tailored to produce the highest product yields and superior culture performance, in CHO cell line systems.

The batch processes performed in this study utilised the same bioreactor conditions as the previous studies, which are described in 3.3. To monitor the progress of the batches, pH, dissolved oxygen and temperature measurements were performed *in-situ* and cell density/viability was done at-line. The metabolite data required for the production of process models was collected in two formats, the offline reference data (YSI: 3.9) and the at-line spectroscopy data (NIR & Raman: 3.5/3.7). Offline measurements were analysed using the techniques and procedures outlined in 3.9, while at-line data was collected from the NIR and Raman spectrometers. The techniques described in 2.3/2.4 were utilised in the production of the final metabolite models for the batch processes.

6.1.1 Aim & Objectives

The aim of this study was to utilise a dual spectroscopic strategy of at-line NIR and Raman spectroscopy, to characterise low passage number batch CHO cell culture in the commercial Dynamis medium. Glucose and lactate were monitored using NIR and Raman for the purpose of developing process models, which would be compared to the CD CHO and CD OptiCHO media processes (Chapter 4 & 5). The information obtained from the overall culture performance and process models, would then be provided to the industrial collaborators which would enable them in further developing an early stage process medium.

The Dynamis processes in this study were performed in the same manner as the CD and CD OptiCHO batch process, so as to provide a suitable comparison. As there are no previous spectral data from either the industrial collaborator or in the published literature, finding a direct comparison upon which to determine the feasibility of the dual spectroscopic strategy in this study has not been straightforward. However an indirect assessment of the strategy, based upon the observations in the previous studies (Chapter 4 & 5) and related studies in the literature, will also be utilised in assessing the spectroscopic strategy in the present study.

As with the previous studies, a combination of pure media analysis and evidence in the published literature were utilised in the determination of the specific regions of the NIR and Raman spectra, throughout the spectral processing of the Dynamis batch samples (Williams and Fleming, 1995, Colthup, 2012, Nyquist and Kagel, 2012, Smith and Dent, 2013). A reassessment of the applied at-line NIR and Raman spectrometers would also be performed to determine if there was any need to optimise the instruments for the analysis of the Dynamis medium process samples. NIR data combined with the YSI reference data

would be utilised to produce quantitative PLS models for both glucose and lactate, while the Raman would be used in a qualitative fashion for these metabolite. Any models produced in the course of the present study would undergo both an internal and external validation, to determine their ability to accurately predict the metabolites.

The research in this study represents a significant final stage of process development, as it utilises the most complex chemically defined medium provided by Thermo Fisher Scientific. Combined with the results obtained from the previous studies, the results of this study will be will be used to provide an insight into the requirements of early stage process development medium.

6.1.2 Novelty

The research presented in the current study adds additional complexity and novelty to that of the previously investigated CHO bioprocesses, the results of which can be fed back to the industrial collaborators. This study adds an additional layer into the spectroscopic investigation and characterisation of low passage number cultures, by testing a batch culture in the alternative chemically defined medium, Dynamis. Secondly these processes will also provide an insight into the performance of Dynamis, which at the time the study was conducted had no published reports in the literature. The level of interest of both industrial collaborators was high indicating the relevance of the study to both Thermo and Lonza.

The interpretation of the spectral data might prove to be more problematic than usual, as there have been no reports of spectroscopic analysis of Dynamis medium based cultures with which to assess any potential spectroscopic variance effects. Similar to the previous study (Chapter 4) to determine if the change to Dynamis medium has a significant

effect on the selection of the optimal data pre-processing strategy the novel DoE strategy was utilised in along with PCA.

6.2 Results

6.2.1 Spectral Region Assignment

To assess whether using an alternative medium to the previous studies influences the presence or location of the glucose and lactate peaks in the spectra of both techniques, samples of the Dynamis media were analysed. Based upon the results, there was no notable effect upon the metabolite signals previously noted in earlier chapters of the current work. When compared to observations from the literature, regions for the NIR and Raman spectra previously assigned to glucose/lactate were utilised for the process sample analysis (Table 6.1) (Chen *et al.*, 2004, Goodarzi *et al.*, 2015, Mehdizadeh *et al.*, 2015, Singh *et al.*, 2015).

Table 6.1: Glucose & lactate peak regions in Dynamis medium.

Metabolite	NIR Peak Region	Raman Peak Region
Glucose	2200-2400 nm	1072 cm^{-1} and 1128 cm^{-1}
Lactate		855 cm^{-1}

6.2.2 Dynamis Batch Process Analysis

6.2.2.1 Reference Analysis

To characterise the performance of CHO 42 cell culture in Dynamis medium, two batch processes were performed as the final process development stage. Characterisation of the process was performed by analysing the live cell density/viability and the cell metabolism through observation of glucose utilisation/lactate production (Figure 6.1) and hIgG production. On day 7 the highest average cell density was observed to be, $1.9x10^6$ cells/mL, which was associated with a steady decrease in cell viability and rapid increase in lactate concentration. One notable feature present in the data is the increase in both cell density and viability on day 12, which coincided with a decrease in lactate concentration. The culture process was terminated on day 14 as a result of an overnight technical fault, however due to the rapidly declining cell density and viability it is unlikely that the processes would have continued for longer. hIgG production has been normalised due to commercial sensitivity, again the notable increase in concentration coincides with the increase in cell density. The spike in hIgG concentration on day 13 also may be related to the increase in cell density/viability on day 12. The decrease in hIgG at the end of the process is most likely the result of degradation brought about by the conditions of the culture. See appendix Table A.5 for the corresponding values to Figure 6.1 and appendix Figure A.5 the bioreactor process condition values (pH, dO_2, Agitation & Temperature).

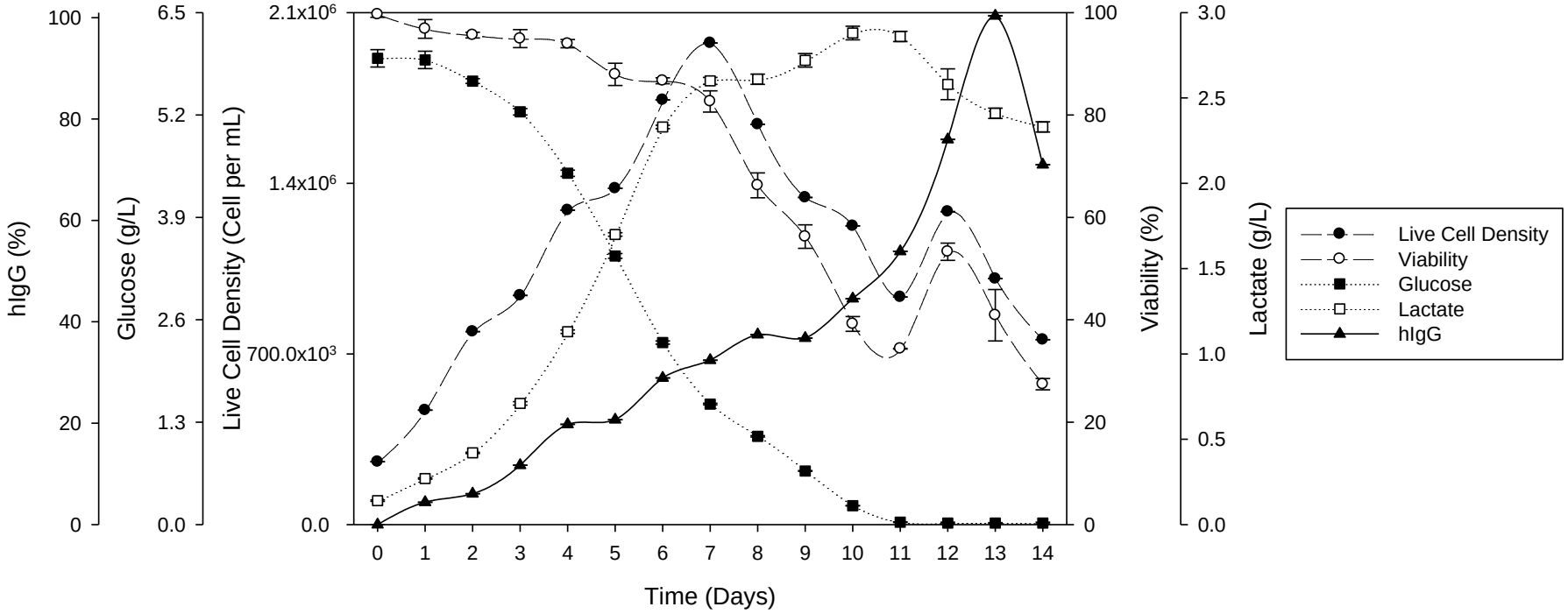


Figure 6.1: Time course of process data from CHO 42 culture in batch Dynamis medium [average of triplicate measurements from two bioreactors plotted with standard deviation].

6.2.2.2 Raw Process Spectra

Due to the effect that water absorbance has upon the NIR spectra, which has been reported in the literature (Beutel and Henkel, 2011, Reich, 2016), no metabolite absorbances which may be present are easily discernible in the unprocessed spectra (Figure 6.2).

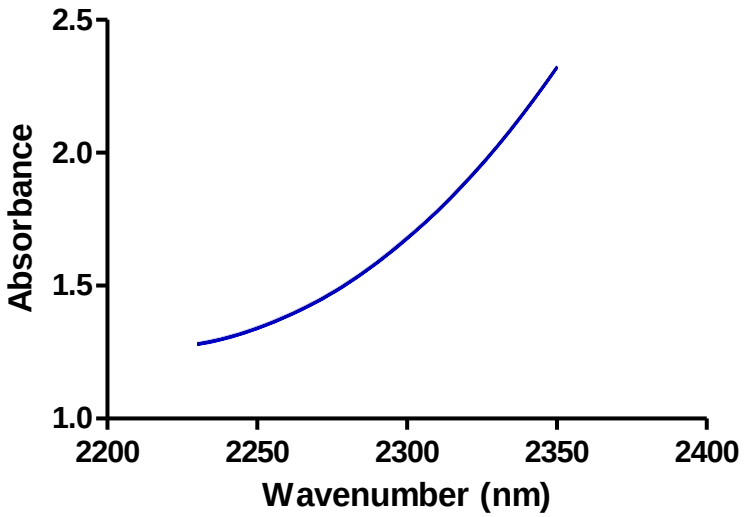


Figure 6.2: Batch Dynamis NIR absorbance of combination band region (stretching and bending vibrations for C–H) from all processes [Colour Code Light to Dark = Bioreactor 1 to 2] (Processing: Average of triplicate measurements from each bioreactor).

The Raman spectral data produced very low intensity peaks at the previously identified regions from the literature, that corresponded to lactate (855 cm^{-1}) and glucose (1128 cm^{-1}) (Mehdizadeh *et al.*, 2015, Singh *et al.*, 2015). For the DoEman and PCA analysis the regions at 835-865 cm^{-1} and 1100-1150 cm^{-1} were chosen to develop the process models, as they represented the regions of highest spectral activity in regards to the target metabolites (Figure 6.3).

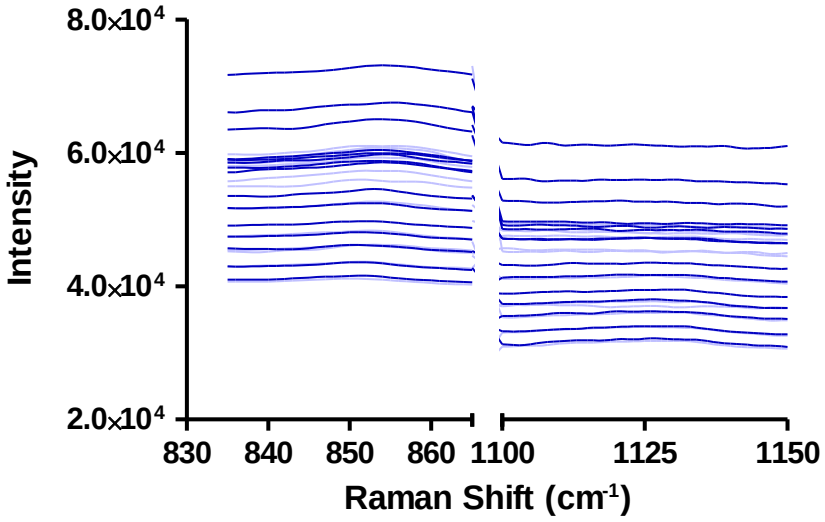


Figure 6.3: Batch Dynamis Raman scattering of 835-865 cm^{-1} and 1100-1150 cm^{-1} regions from all processes [Colour Code Light to Dark = Bioreactor 1 to 2] (Processing: Average of triplicate measurements from each bioreactor).

6.2.2.3 DoEman

Working under the assumption that a change in medium represents a significant change in the sample matrix, and could change the previously identified pre-processing techniques, a new DoE analysis was performed for the Dynamis medium. The DoE strategy previously outlined in 3.8.2 was applied with the selected pre-processing treatments indicated in Table 6.2, to determine the optimal data processing strategy (Table 3.2

6.2.2.4 PCA Scores/Plots

The second step in the spectral data optimisation was a PCA of the previously identified regions of the NIR and Raman spectra. The PCA parameters which captured the maximum variance within both the NIR and Raman data are outlined in Table 6.3. From these results are the calibration and validation datasets were selected.

Table 6.3: Spectral pre-processing PCA parameters and statistical values

Technique	Pre-processing	Cross Validation	No. of Latent Variables	Variance Captured
NIR	2nd Derivative, MSC & Mean Centering	Venetian Blinds	4	98.72 %
Raman	MSC & Mean Centering	Venetian Blinds	4	93.44 %

The NIR PCA strategy outlined in Table 6.3 resulted in the PCA scores plot observed in Figure 6.4 and despite the apparent difference between the two batches there is a distinct trend between PC1 vs PC2 shared by both batches. In both batches, early process samples correlate to PC1 and then favour PC2 as the process time increases. From these results, the calibration dataset was formed from two thirds of the data which covered the PCA scores plot, while the remaining third formed the external validation data set. This represents a rigorous external validation of the model formulated using the calibration dataset.

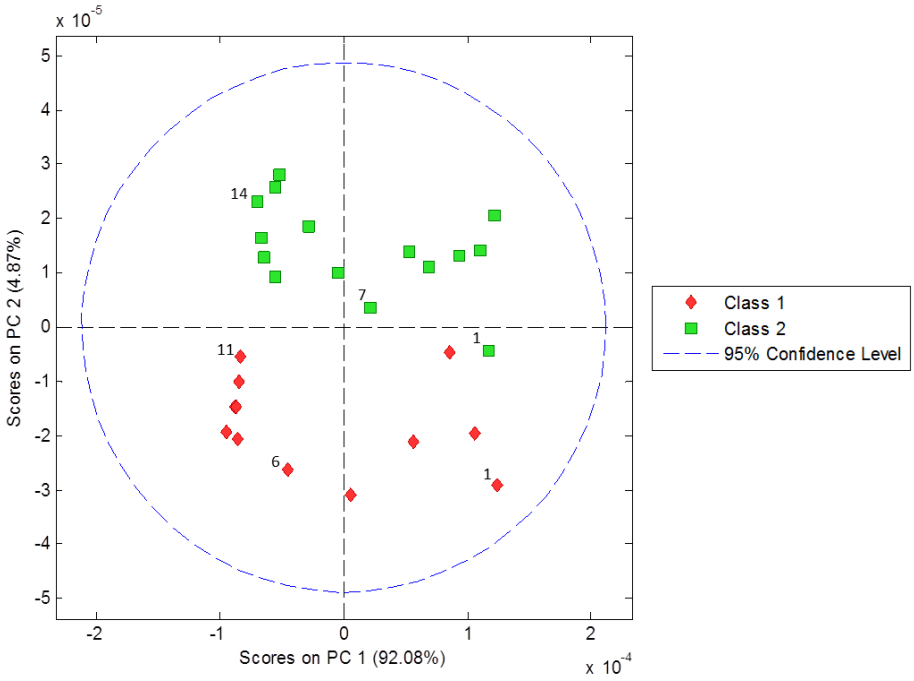


Figure 6.4: Batch Dynamis NIR PCA scores of the combination band region in all processes, with process sample numbers represented [Class=Bioreactor 1-2] (Processing: Savitzky-Golay Second Derivative; MSC; Mean Centered; Average of triplicate measurements per batch).

The Raman PCA results (Figure 6.5) also displayed the previously observed trend from the NIR, however there is more overlap between the batches compared to that observed in the NIR which indicate greater similarity in the composition of those samples. As before the early samples are correlated to PC1 and the later process samples develop a correlation to PC2. The calibration dataset was again composed of two thirds of the data and the external validation dataset was produced from the other third.

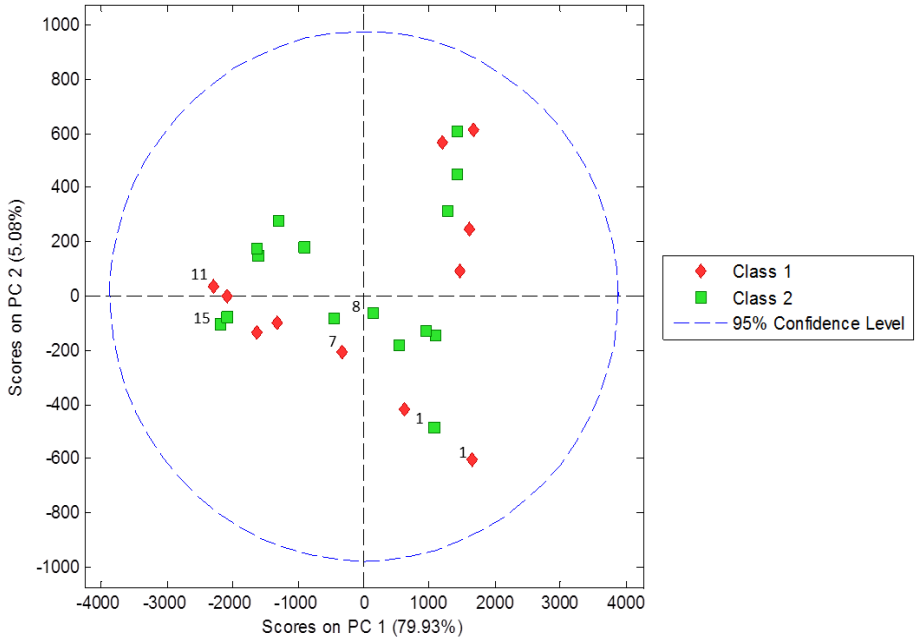


Figure 6.5: Batch Dynamis Raman PCA scores of the 835-865 cm^{-1} and 1100-1150 cm^{-1} regions from all processes, with process sample numbers represented [Class=Bioreactor 1-2] (Processing: MSC and Mean Centered; Average of triplicate measurements per batch).

Despite the apparent variance between the batches in both the NIR and Raman, the trend observed within the PCA scores plots of each technique indicated that PC1 and PC2 correlate to glucose and lactate, respectively.

6.2.2.5 Processed Spectra

After the application of the identified optimal pre-processing techniques to the raw NIR data, peaks and troughs corresponding to the absorbance intensity of the CH region are now observable (Figure 6.6). The varying concentration of glucose/lactate in the reference data over the process time is reflected in the changes in height and depth of the peaks from both batches.

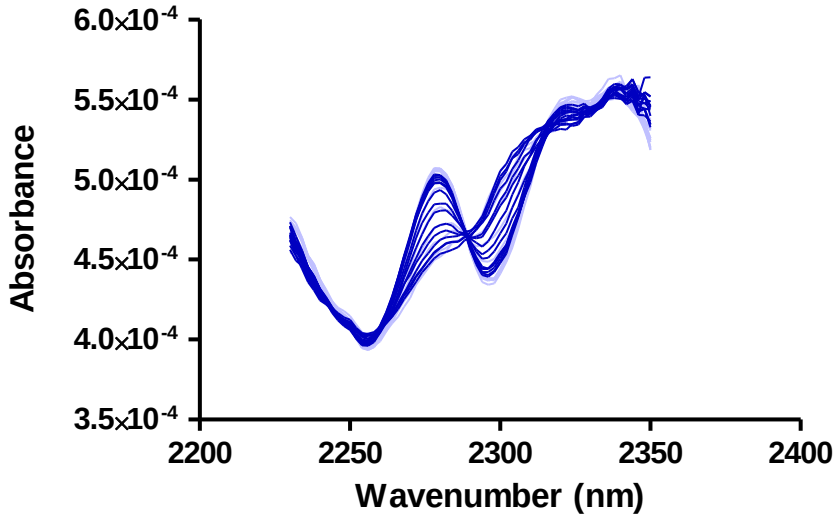


Figure 6.6: Preprocessed batch Dynamis NIR absorbance of CH region of all processes [Colour Code Light to Dark = Bioreactor 1 to 2] (Processing: Savitzky-Golay Second Derivative; MSC; Mean Centered; Average of triplicate measurements from each bioreactor).

The previously observed peak activity at the target spectral regions are better elucidated in the Raman spectral data after the application of the optimised pre-processing treatments (Figure 6.7). Again based upon the reference data, the intensity of the peaks at the specified regions is representative of the concentration of glucose and lactate within the samples of both batches.

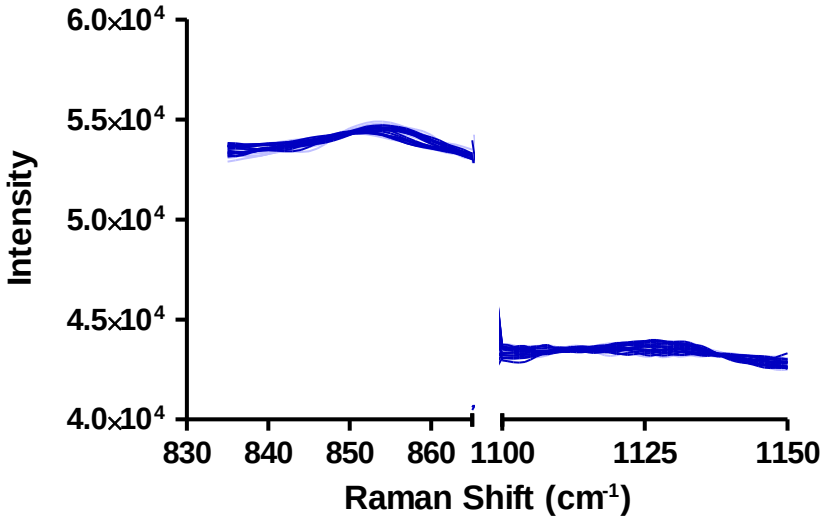


Figure 6.7: Processed batch Dynamis Raman scattering of 835-865 cm^{-1} and 1100-1150 cm^{-1} regions from all processes [Colour Code Light to Dark = Bioreactor 1 to 2] (Processing: MSC; Mean Centered; Average of triplicate measurements from each bioreactor).

6.2.2.6 Partial Least Squares (Calibration & Validation)

Once the spectral data optimisation was completed and the calibration/validation datasets selected, the final stage of the process characterisation was to carry out the PLS modelling of the target metabolites. The NIR data produced PLS models which indicated a good model fit and with low errors of prediction both glucose (Figure 6.8) and lactate (Figure 6.9). However there is a slight offset in the glucose model which is not present in the lactate model.

[A]

	Calibration	Internal Validation	External Validation (Prediction)
RMSE (g/L)	0.23	0.52	0.63
R^2	0.99	0.96	0.94

[B] **[C]**

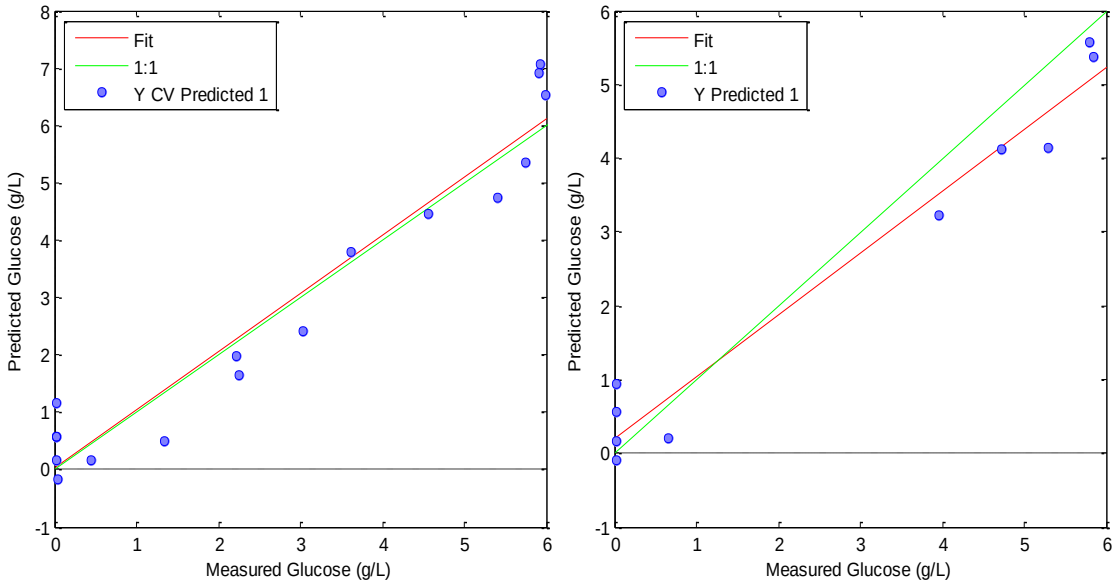


Figure 6.8: NIR Batch Dynamis glucose PLS [A] RMSE & R^2 values [B] Calibration & Internal Validation model [C] External Validation (Prediction) model; with the selected samples (blue points), the theoretical 'perfect' fit (green line) and the actual fit (red line).

[A]

	Calibration	Internal Validation	External Validation (Prediction)
RMSE (g/L)	0.04	0.13	0.08
R^2	0.99	0.99	0.99

[B] **[C]**

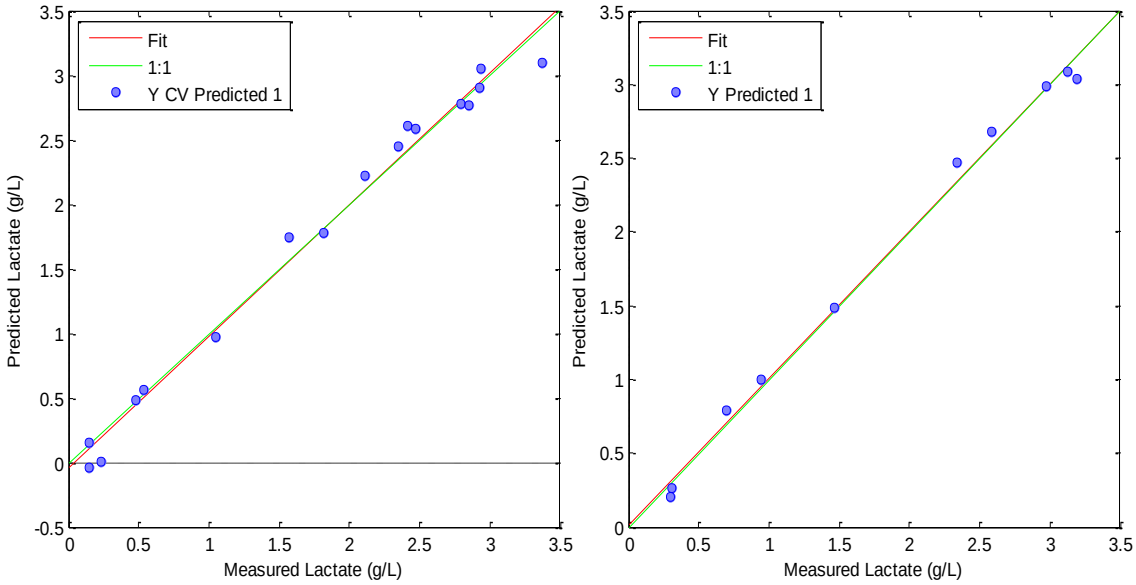


Figure 6.9: NIR Batch Dynamis lactate PLS [A] RMSE & R^2 values [B] Calibration & Internal Validation model [C] External Validation (Prediction) model; with the selected samples (blue points), the theoretical 'perfect' fit (green line) and the actual fit (red line).

Once more the Raman data produced poor PLS models with high errors of prediction for both glucose (Figure 6.10) and lactate (Figure 6.11). Again the poor predictive abilities of the Raman PLS were not unexpected, based on the observations from the previous studies (Chapter 4 & 5). Despite the qualitative success of the Raman it is still not clear what the cause of the poor model performance is and if it is the result of undetected effects from other sample matrix components or poor selection of calibration/validation datasets.

[A]

	Calibration	Internal Validation	External Validation (Prediction)
RMSE (g/L)	0.34	3.02	3.16
R^2	0.98	0.01	0.16

[B] **[C]**

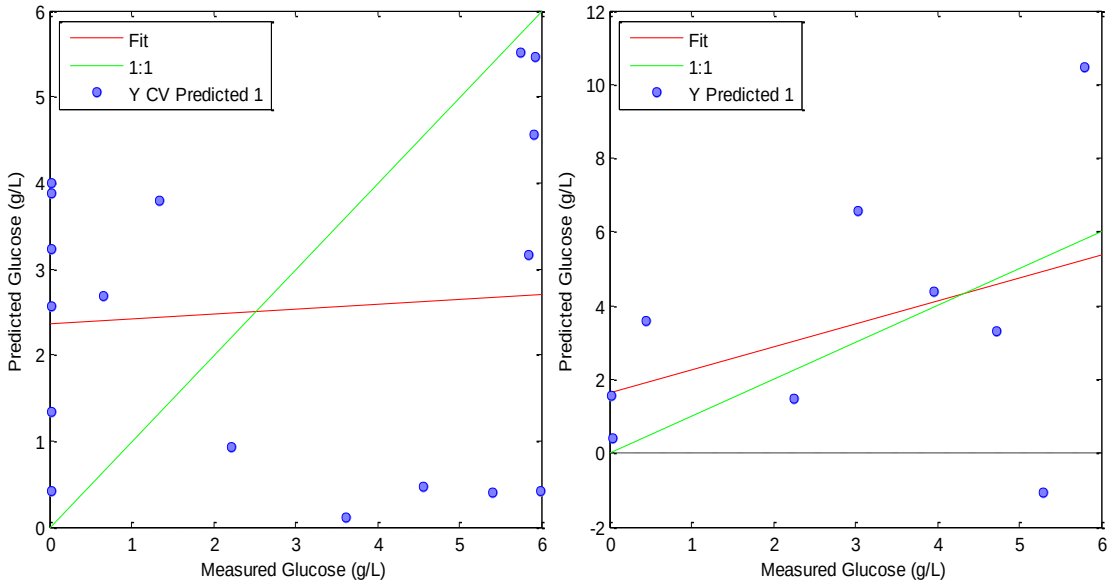


Figure 6.10: Raman Batch Dynamis glucose PLS [A] RMSE & R^2 values [B] Calibration & Internal Validation model [C] External Validation (Prediction) model; with the selected samples (blue points), the theoretical 'perfect' fit (green line) and the actual fit (red line).

[A]

	Calibration	Internal Validation	External Validation (Prediction)
RMSE (g/L)	0.16	1.41	1.35
R^2	0.98	0.01	0.17

[B] **[C]**

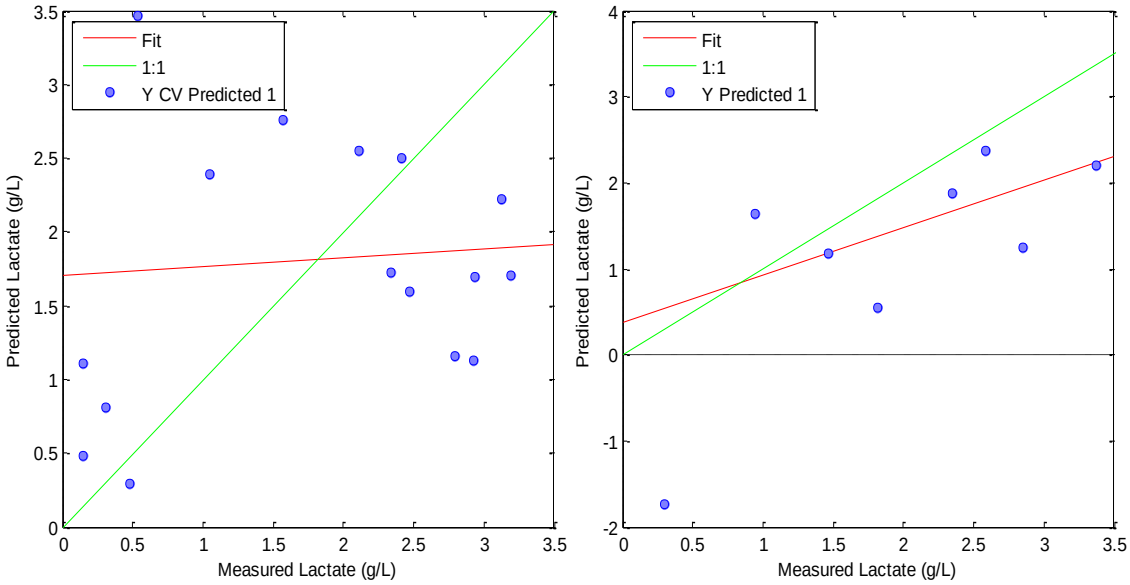


Figure 6.11: Raman Batch Dynamis lactate PLS [A] RMSE & R^2 values [B] Calibration & Internal Validation model [C] External Validation (Prediction) model; with the selected samples (blue points), the theoretical 'perfect' fit (green line) and the actual fit (red line).

The results produced from the PLS modelling demonstrate the suitability of NIR for predicting glucose and lactate, in the Dynamis CHO culture. But again indicate poor suitability of the Raman models in predicting both of the metabolites.

6.3 Discussion

6.3.1 Spectral Models

6.3.1.1 NIR

The glucose error of prediction for the external validation obtained in this study was 0.11 g/L greater than that of the internal validation, indicating an over fit in the initial calibration. This observed over fit of the initial calibration was absent from the lactate data, with the error of prediction for the internal validation being 0.05 g/L than those of the external validations. This difference in the errors of prediction between the two metabolites could potentially be a result of the sample distribution for the calibration/validation datasets favouring lactate over glucose. This is reflected in the R^2 values of the external validation for the metabolites, with glucose being 0.94 and lactate 0.99, indicating a better correlation between the spectral data to the lactate reference data.

In terms of drawing a direct comparison between the glucose/lactate results produced in the present study to the published literature has not been possible. This is due to there being no published sources which have utilised Dynamis medium or tried to apply NIR to such a process. An indirect comparison of NIR in Dynamis can be made from the previous CHO media batch processes (4.3.1.1 & 5.3.1.1) and those in the published literature. Although the glucose model error in the Dynamis batches, 0.62 g/L, is higher than those obtained from the CD and OptiCHO media (0.45 g/L and 0.37 g/L), they are still within the range of those reported in some of the literature 0.48-1.45 g/L (Lee *et al.*, 2012a, Sandor *et al.*, 2013, Milligan *et al.*, 2014). Conversely, the lactate results produced in the Dynamis batch (0.08 g/L) were lower than those obtained in both the, previous studies,

0.34/0.17 g/L for the CD CHO/CD OptiCHO respectively, and in the CD CHO literature, 0.44 g/L (Sandor *et al.*, 2013) .

Overall the results from this study indicate an acceptable level of error within the models for both metabolites, glucose (0.62 g/L) and lactate (0.08 g/L), which compare well to the previous studies and the published CHO culture literature. These low errors are potentially useful to the cell culture industry, as they indicate the suitability to use NIR to monitor an industrial process with an acceptable degree of accuracy.

6.3.1.2 Raman

Once more there was no change in the quantitative capabilities of the Raman, despite the spectral signals in the previously identified regions, for glucose (1128 cm^{-1}) and lactate (855 cm^{-1}) (Mehdizadeh *et al.*, 2015, Singh *et al.*, 2015). Regardless of the application of the developments outlined in 4.3.1.3, the Raman data has failed to produce quantitative models for both metabolites in this study.

Similar to the NIR, as there is no application of Dynamis processes in the literature gaining a direct comparison of the results from the Raman spectroscopic data was not possible. The results indicate the models in their current state would not be suitable for use in monitoring an industrial process, as they are not sufficiently accurate at predicting the metabolite concentrations, which is indicated by the RMSEP values 3.16 g/L (glucose) and 1.35 g/L (Lactate).

When indirectly compared to the previous studies, the results generated here have similarly high errors of prediction to those of the previous media batches (4.3.1.1 & 5.3.1.1). Both the glucose/lactate model errors in the Dynamis medium are higher than those observed from the previous CD CHO and CD OptiCHO studies, 3.11/1.62 g/L and

2.04/0.93 g/L. When compared to the glucose and lactate results obtained from one CD OptiCHO study, 0.27 g/L and 0.20 g/L (Matthews et al., 2016), it is clear that the errors here are significantly higher and the present models are not suitable for process monitoring at this time without more development and optimisation.

6.3.2 Batch Process

In terms of meeting the needs of the industrial collaborators, Lonza and Thermo Fisher Scientific, this low passage number batch process represents the final characterisation stage, as the Dynamis medium is the most recently developed chemically defined medium.

From the reference results it is clear that the performance of the culture is limited, as indicated by the slow growth rate and decreasing viability, which are potentially related to the lactate increasing concentration. This assumption is supported by the increase of both the cell density and viability on day 12, which coincided with a decrease in lactate concentration.

The decrease in the lactate concentration indicates a potential metabolic switch to lactate into the TCA cycle as an alternate energy source around day 11, as this is when the glucose concentration is almost completely depleted. As previously reported in the literature, one of the causes of this switch is reduced glucose availability (Tsao *et al.*, 2005, Wilkens *et al.*, 2011, Zagari *et al.*, 2013), which could be applied to the batch process presented in this study.

As previously stated in 6.2.2.1, the processes had to be terminated on day 14 due to a technical fault. But regardless of this unfortunate circumstance, it is clear from the data that both the cell density and viability were on the decline again, despite the potential switch to lactate consumption to support the culture.

6.3.3 Dynamis vs CD CHO & CD OptiCHO

To determine how the Dynamis medium performed, the results from this study will be compared to those obtained from the previous CD CHO and OptiCHO batch processes only. Despite the additional day in the process time, this has not had any beneficial effect upon the culture. Comparing the highest cell densities, the result of the Dynamis is very low at only $1.9x10^6$ cells/mL, almost a third lower that of the CD CHO medium ($5.8x10^6$ cells/mL) and lower still than the CD OptiCHO ($2.5x10^6$ cells/mL).The low viable cell density in the Dynamis medium, has subsequently also resulted in a lower yields of hIgG being attained, composed to those from the CD CHO and CD OptiCHO media. In terms of metabolic profiles from all three studies they are relatively comparable, with glucose being reduced below 1 g/L from days 7-9 and lactate reaching the highest concentration of 2.6-2.9 g/L between days 8-10.

A number of observations can be made about the performance of the cell line across each medium, which can be fed back to the industrial collaborators to help in the development of early culture process medium. The similarity between the culture time and the metabolic profile (glucose consumption/lactate production) across each medium process indicates that the basic metabolic needs of the cell line are being met fairly consistently in each medium. However, the changes in cell density between each medium process indicate that of the three CD CHO produces the best results. In terms of the highest average hIgG concentration between the three media, the Dynamis batch process performed only marginally better than the CD OptiCHO media, producing half as much as the CD CHO batch process. Again, the results of the three studies indicate that the CD CHO medium is better suited to early culture development, due to the increased viable cell density and subsequent hIgG production. But with the specific formulation of the medium

being proprietary, it is not possible to specifically identify which components are responsible for the varying performances of the cultures.

6.4 Conclusion

The present study once again clearly demonstrates the qualitative and quantitative capabilities of NIR spectroscopy in the monitoring of mammalian cell culture in batch processes, and the qualitative ability of Raman in the analysis of glucose and lactate. The study also provides a suitable starting point for further investigation into the utilisation of Dynamis medium.

The NIR batch and fed batch processes produced successful models for glucose and lactate, with acceptable RMSE and R^2 values. Although there is no reporting of NIR in Dynamis medium, the results obtained were comparable to those in the published literature in CD CHO and CD OptiCHO media. Again despite the Raman data producing qualitative signals in the glucose and lactate target regions of the spectra, this did not result in successful modelling of the data and the observed signals were weaker in intensity that the previous two studies.

In conclusion this study indicates that in comparison to the CD CHO and CD OptiCHO media, Dynamis is not as suitable for low passage cultures, as it produces lower viable cell densities and hIgG concentration. This could possibly be remedied by optimisation of the culture work up and further adaptation to the “new” medium of the cell line. In the present study, time limitations prevented this.

Chapter 7:

Conclusions

7.1 Conclusions

The main aim at the outset of this project was to look at approaches for progressing towards real-time monitoring and control of mammalian cell line based biomanufacturing processes. To achieve this, NIR, MIR and Raman were utilised at-line to construct metabolite models in a number of industrial CHO cell culture processes. This project investigated the performance of low passage number cultures in three different media, to align with the needs of the industrial collaborators, and utilised a DoE strategy to optimise spectral data preprocessing when forming models for each of the cell culture process. The final aspect that was investigated was the formulation of a single model from the data of all the spectroscopic techniques through data fusion. An overview of the processes and analysis performed throughout this study, to achieve these aims, can be seen in Table 8.1.

Table 7.1: Study experimental overview

Chapter	Medium	Process Format	Spectroscopic Analysis		
4	CD CHO	Batch & Fed Batch	NIR	MIR	Raman
5	CD Opti CHO		NIR & Raman		
6	Dynamis	Batch	NIR & Raman		

The results presented from the initial study (Chapter 4) in the CD CHO medium process demonstrated varying degrees of success between the spectroscopic techniques and the process formats. Utilising the combination band region in the NIR spectra, glucose and lactate models were successfully constructed in both the batch and fed batch process formats, which compared well with the literature. In contrast the MIR and Raman both failed to produce process models without high errors, despite the latter displaying qualitative Raman signals for the target metabolites. While the CD CHO batch process

formed a suitable baseline for the characterisation of the low passage number cell cultures, the unrefined fed batch protocol used adversely affected the overall process. Disappointingly, as only the NIR produced suitable process models, investigating the feasibility of data fusion was not possible. This initial study set a good foundation in terms of the low passage number batch culture characterisation and NIR analysis of CD CHO. But it also highlighted the shortcomings of the Raman analysis and the failure of the MIR. As a result of this some of the previously planned analysis had to be reassessed, with only NIR and Raman being carried forth in the investigation of CD OptiCHO processes.

The outcomes of the CD OptiCHO study (Chapter 5) study reinforced the suitability of NIR and Raman as quantitative/qualitative techniques respectively, for glucose/lactate analysis in both batch/fed batch processes. Despite limited application of NIR in CD OptiCHO medium, there were indications that the results in the present study were better than those few published studies. Modelling of the Raman data was again unsuccessful in producing process models, despite the presence of qualitative signals in the target regions. Likewise the fed batch strategy used in this study proved to be significantly detrimental to the cell culture process, increasing the decline in live cell density leading to a halving of the overall process time. Compared to the CD CHO medium processes, the results of the CD OptiCHO medium processes indicate it is unsuitable for the development of low passage cultures, due to the lower viable cell density and hIgG concentration produced. The results of this study continue to build upon the initial batch process characterisation through the NIR/Raman analysis and highlight the limitations of the more developed CD OptiCHO medium.

The final study in this project (Chapter 6) represented the not only the last stage of process characterisation for the industrial partners, but also a suitable starting point for

further investigation into the utilisation of Dynamis medium. The NIR data from both the batch and fed batch processes produced successful models for glucose and lactate, with satisfactory error values comparable to CD CHO and CD OptiCHO media research in the published literature. Once more the qualitative signals in the glucose and lactate target regions of the Raman spectra data did not yield successful models. When compared to the results from CD CHO and CD OptiCHO, those obtained from the Dynamis indicate that it was the least suitable of the three for use in low passage cell cultures, because of the low viable cell density and hIgG concentration. Overall this study proved to be a suitable finishing point of the investigation of the CHO cell processes, demonstrating again the success of NIR and the poor performance of the Dynamis medium for low passage number cultures.

In conclusion the results achieved in throughout this project demonstrate the success of NIR analysis across all process formats and media, as supported by the literature, and the failure of MIR based data modelling. The qualitative analytical ability of Raman spectroscopy is also shown, with the results of this analysis appearing to differ between the various media processes. Raman spectroscopy using the instrument supplied did not seem suitable for quantitative modelling despite published research indicating otherwise. The research in this project also provides a suitable indicator as to the media suitability for use in low passage number cell cultures, through comparison of the batch culture processes. Overall this study represents a useful stage in the development towards real time monitoring of biomanufacturing processes and in the development of low passage number cell culture media, however it requires further investigation to improve upon the Raman/MIR analysis and fed batch strategy, but also to refine the processes for future use in industry.

Chapter 8:

Future Work

8.1 Future Work

As with all research studies there are always areas which can be improved in hindsight, however the research presented here does offer a foundation from which future developments can be made. But due to both the time limitations and some of the unexpected results, there are a number of aspects which would benefit from future improvements and developments.

8.2 NIR

Although the at-line NIR proved successful in monitoring of glucose and lactate in all process formats, there are still some improvements and developments that can still be made. Investigation into *in-situ* NIR monitoring should be assessed, as it would be the key next stage in development towards real time monitoring of these processes and could potentially reduce some of the limitations observed in this study. The main limitation to the at-line was the ability to analyse other important metabolites (glutamine) and culture parameters (viable cell density), which give a wider understanding of the process. As yet there are no studies which have looked at *in-situ* NIR monitoring in low passage number cultures combined with DoE based preprocessing. But number of studies have indicated that application of *in-situ* NIR can yield process models for multiple culture process parameters (Clavaud *et al.*, 2013, Courtès *et al.*, 2016).

8.3 MIR

As previously discussed the failure of the MIR was potentially caused by a number of reasons (4.3.1.2). There are a couple of methods which could be used to try to improve the MIR analysis in the future. Firstly, moving the instrument to a more suitable location within the processing area would be done, to determine if the poor temperature control in the

current location was to blame for the poor performance. A more extensive study into the limit of detection of the instrument would also be carried out to reassess the suitability of the instrument. The final potential improvement would be to switch to a different instrument such as the Direct Detect at-line MIR spectrophotometer (Merck Millipore, Danvers, MA). A few studies have demonstrated that this instrument, which uses a PTFE membrane to present the sample to the spectrometer, has been successful in analysing various parameters in CHO cell cultures (Capito *et al.*, 2015a, Capito *et al.*, 2015b).

8.4 Raman

The ability for the Raman spectrometer to only produce data of a qualitative data was disappointing, even considering the improvements to the instrument that were made (4.3.1.3). As these improvements indicated no fault with the instrument, the only conceivable development to potentially improve the process spectra would be to increase the analysis time through *in-situ* analysis. There are a number of reported successes in the literature of *in-situ* Raman (Whelan *et al.*, 2012, Matthews *et al.*, 2016) but not within a low passage number process combined with DoE , based pre-processing, thus maintaining the original novel aspects of this thesis.

8.5 Modelling

Overall the modelling strategy employed in this thesis was suitable, as indicated by some of the successfully developed models. However, one improvement that could be made to the modelling would be to increase the number of processes utilised in each developed model. This would increase the amount of spectral measurements potentially leading to better end models.

8.6 Medium

In terms of meeting the aim of characterising low passage culture performance in the various media, there is still a significant amount of development to be done, but the work presented here forms a good foundation for future studies. Looking at the major performance indicators, viable cell density, viability and product titre, it is clear that the CD CHO medium was most suitable for low passage number culture cultivation. As the three media in this study were proprietary and the compositions not disclosed to the author, it is possible that a removal or addition of components to the media is the cause of the poor performances but this can only be speculated. Assuming that it is a removal or addition of components to the media that caused the issues, one method of future development could be to perform multiple experiments in CD CHO batch processes to fully characterise the metabolic needs of low passage cultures. This could then enable the industrial partners involved, Thermo Fisher Scientific, to tailor a medium specifically for low passage number cultures.

8.7 Fed Batch Strategy

The fed batch strategies in this thesis represent one of the major future developments that must be addressed, due to the failure of these in the CD CHO and CD OptiCHO processes. While these strategies were chosen from the literature as they demonstrated an improved culture performance (Barrett *et al.*, 2012, Reinhart *et al.*, 2015a), it could be that these feeds are not compatible with low passage number cultures. Had time permitted, a series of small scale feasibility studies would have been very useful to determine the compatibility of these feeds and an optimisation of the feeding strategy.

Chapter 9:

References

Abdi, H. 2003. Partial least square regression (PLS regression). *Encyclopedia for research methods for the social sciences***,** 792.

Abdi, H. & Williams, L. J. 2010. Principal component analysis. *WIREs Comput Stat,* 2**,** 433.

Abu-Absi, N. R*., et al.* 2011. Real Time Monitoring of Multiple Parameters in Mammalian Cell Culture Bioreactors Using an In-Line Raman Spectroscopy Probe. *Biotechnology and Bioengineering,* 108**,** 1215.

Afseth, N. K. & Kohler, A. 2012. Extended multiplicative signal correction in vibrational spectroscopy, a tutorial. *Chemometrics and Intelligent Laboratory Systems,* 117**,** 92.

Ashton, L. & Goodacre, R. 2011. Application of Deep UV Resonance Raman Spectroscopy to Bioprocessing. *European Pharmaceutical Review,* 16.

Ashton, L*., et al.* 2013. The challenge of applying Raman spectroscopy to monitor recombinant antibody production. *Analyst,* 138**,** 6977.

Bachmann, C*., et al.* 2013. A comparative assessment of multi-sensor data fusion techniques for freeway traffic speed estimation using microsimulation modeling. *Transportation research part C: emerging technologies,* 26**,** 33.

Baik, J. Y*., et al.* 2015. Optimization of bioprocess conditions improves production of a CHO cell-derived, bioengineered heparin. *Biotechnology journal,* 10**,** 1067.

Baik, J. Y*., et al.* 2012. Metabolic engineering of Chinese hamster ovary cells: towards a bioengineered heparin. *Metabolic engineering,* 14**,** 81.

Bailey, L. A*., et al.* 2012. Determination of Chinese hamster ovary cell line stability and recombinant antibody expression during long-term culture. *Biotechnology and bioengineering,* 109**,** 2093.

Bandaranayake, A. D. & Almo, S. C. 2014. Recent advances in mammalian protein production. *FEBS letters,* 588**,** 253.

Barrett, S. L*., et al.* 2012. Attaining Next-Level Titers In CHO Fed-Batch Cultures. *BioProcess International,* 10**,** 10.

Becker, T*., et al.* 2007. Future aspects of bioprocess monitoring. *In:* ULBER, R. & SELL, D. (eds.) *White Biotechnology.*

Beckmann, T. F*., et al.* 2012. Effects of high passage cultivation on CHO cells: a global analysis. *Applied Microbiology and Biotechnology,* 94**,** 659.

Berry, B*., et al.* 2015. Cross-scale predictive modeling of CHO cell culture growth and metabolites using Raman spectroscopy and multivariate analysis. *Biotechnology progress,* 31**,** 566.

Beuermann, T*., et al.* 2012. On-line carbon balance of yeast fermentations using miniaturized optical sensors. *Journal of Bioscience and Bioengineering,* 113**,** 399.

Beutel, S. & Henkel, S. 2011. In situ sensor techniques in modern bioprocess monitoring. *Applied Microbiology and Biotechnology,* 91**,** 1493.

Biancolillo, A*., et al.* 2014. Data-fusion for multiplatform characterization of an Italian craft beer aimed at its authentication. *Anal Chim Acta,* 820**,** 23.

Biechele, P*., et al.* 2015. Sensor systems for bioprocess monitoring. *Engineering in Life Sciences,* 15**,** 469.

Brereton, R. G. 2007. *Applied chemometrics for scientists*, John Wiley & Sons.

Buss, N. A*., et al.* 2012. Monoclonal antibody therapeutics: history and future. *Current opinion in pharmacology,* 12**,** 615.

Capito, F*., et al.* 2015a. At-line mid infrared spectroscopy for monitoring downstream processing unit operations. *Process Biochemistry,* 50**,** 997.

Capito, F*., et al.* 2013. Host cell protein quantification by fourier transform mid infrared spectroscopy (FT-MIR). *Biotechnology and Bioengineering,* 110**,** 252.

Capito, F*., et al.* 2015b. Mid-infrared spectroscopy-based analysis of mammalian cell culture Parameters. *Biotechnology progress,* 31**,** 578.

Challa, S. & Potumarthi, R. 2012. Chemometrics-Based Process Analytical Technology (PAT) Tools: Applications and Adaptation in Pharmaceutical and Biopharmaceutical Industries. *Applied Biochemistry and Biotechnology*.

Chen, J*., et al.* 2004. Comparison of combination and first overtone spectral regions for near-infrared calibration models for glucose and other biomolecules in aqueous solutions. *Analytical chemistry,* 76**,** 5405.

Chen, Q*., et al.* 2015. Classification of different varieties of Oolong tea using novel artificial sensing tools and data fusion. *LWT-Food Science and Technology,* 60**,** 781.

Clavaud, M*., et al.* 2013. Chemometrics and in-line near infrared spectroscopic monitoring of a biopharmaceutical Chinese hamster ovary cell culture: prediction of multiple cultivation variables. *Talanta,* 111**,** 28.

Clementschitsch, F. & Bayer, K. 2006. Improvement of bioprocess monitoring: development of novel concepts. *Microbial Cell Factories,* 5.

Colthup, N. 2012. *Introduction to infrared and Raman spectroscopy*, Elsevier.

Courtès, F*., et al.* 2016. A dual near-infrared and dielectric spectroscopies strategy to monitor populations of Chinese hamster ovary cells in bioreactor. *Biotechnology letters,* 38**,** 745.

Datta, P*., et al.* 2013. An 'omics approach towards CHO cell engineering. *Biotechnology and Bioengineering*.

De Beer, T. R. M*., et al.* 2008. Raman spectroscopy as a process analytical technology (PAT) tool for the in-line monitoring and understanding of a powder blending process. *Journal of Pharmaceutical and Biomedical Analysis,* 48**,** 772.

Dean, J. & Reddy, P. 2013. Metabolic analysis of antibody producing CHO cells in fed batch production. *Biotechnology and Bioengineering*.

Dearing, T. I*., et al.* 2011. Characterization of Crude Oil Products Using Data Fusion of Process Raman, Infrared, and Nuclear Magnetic Resonance (NMR) Spectra. *Applied Spectroscopy,* 65**,** 181.

Del Val, I. J*., et al.* 2012. Application of Quality by Design Paradigm to the Manufacture of Protein Therapeutics. *In:* PETRESCU, S. (ed.) *Glycosylation.* InTech.

Ecker, D. M*., et al.* The therapeutic monoclonal antibody market. MAbs, 2015. Taylor & Francis, 9.

F.D.A. 2004. PAT—A framework for innovative pharmaceutical development, manufacturing, and quality assurance. *In:* F.D.A. (ed.).

Faber, N. & Rajko, R. 2007. How to avoid over-fitting in multivariate calibration—The conventional validation approach and an alternative. *Analytica Chimica Acta,* 595**,** 98.

Farid, S. S. 2007. Process economics of industrial monoclonal antibody manufacture. *Journal of Chromatography B,* 848**,** 8.

Flåten, G. R. & Walmsley, A. D. 2003. Using design of experiments to select optimum calibration model parameters. *Analyst,* 128**,** 935.

Foley, R*., et al.* 2012. Potential of Mid-Infrared Spectroscopy for On-Line Monitoring of Mammalian Cell Culture Medium Components. *Applied Spectroscopy,* 66**,** 33.

Forshed, J*., et al.* 2007. Evaluation of different techniques for data fusion of LC/MS and 1H-NMR. *Chemometrics and Intelligent Laboratory Systems,* 85**,** 102.

Gagnon, M*., et al.* 2011. High-end pH-controlled delivery of glucose effectively suppresses lactate accumulation in CHO fed-batch cultures. *Biotechnology and Bioengineering,* 108**,** 1328.

Galluzzi, L.*, et al.* 2012. Trial Watch: Monoclonal antibodies in cancer therapy. *Oncoimmunology,* 1**,** 28.

Geladi, P.*, et al.* 1985. Linearization and scatter-correction for near-infrared reflectance spectra of meat. *Applied spectroscopy,* 39**,** 491.

Gemperline, P. 2006a. *Practical Guide To Chemometrics*.

Gemperline, P. 2006b. Practical Guide to Chemometrics (2nd). Taylor & Francis Group, LLC: Boca Raton.

Genkawa, T.*, et al.* 2012. Development of a Near-Infrared/Mid-Infrared Dual-Region Spectrometer for Online Process Analysis. *Applied Spectroscopy,* 66**,** 773.

Geörg, D.*, et al.* 2015. MIR-ATR sensor for process monitoring. *Measurement Science and Technology,* 26**,** 065501.

Gerretzen, J.*, et al.* 2016. Boosting model performance and interpretation by entangling preprocessing selection and variable selection. *Analytica Chimica Acta,* 938**,** 44.

Gerretzen, J.*, et al.* 2015. Simple and Effective Way for Data Preprocessing Selection Based on Design of Experiments. *Analytical chemistry,* 87**,** 12096.

Glassey, J. 2016. Multivariate Modeling for Bioreactor Monitoring and Control. *Bioreactors: Design, Operation and Novel Applications*.

Glassy, M. C. & Gupta, R. 2014. Technical and ethical limitations in making human monoclonal antibodies (An Overview). *Human Monoclonal Antibodies.* Springer.

Gomes, J.*, et al.* 2015. Integrating systems analysis and control for implementing process analytical technology in bioprocess development. *Journal of Chemical Technology and Biotechnology,* 90**,** 583.

Goodarzi, M.*, et al.* 2015. Multivariate calibration of NIR spectroscopic sensors for continuous glucose monitoring. *TrAC Trends in Analytical Chemistry,* 67**,** 147.

Gowen, A.*, et al.* 2011. Preventing over-fitting in PLS calibration models of near-infrared (NIR) spectroscopy data using regression coefficients. *Journal of Chemometrics,* 25**,** 375.

Grassi, S.*, et al.* 2014. Beer fermentation: monitoring of process parameters by FT-NIR and multivariate data analysis. *Food chemistry,* 155**,** 279.

Guo, Q.*, et al.* 1999. The robust normal variate transform for pattern recognition with near-infrared data. *Analytica chimica acta,* 382**,** 87.

Hakemeyer, C.*, et al.* 2012. At-line NIR spectroscopy as effective PAT monitoring technique in Mab cultivations during process development and manufacturing. *Talanta,* 90**,** 12.

Han, Y. K.*, et al.* 2011. Autophagy and apoptosis of recombinant Chinese hamster ovary cells during fed-batch culture: Effect of nutrient supplementation. *Biotechnology and bioengineering,* 108**,** 2182.

Harms, P.*, et al.* 2002. Bioprocess monitoring. *Current Opinion in Biotechnology,* 13**,** 124.

Huang, Y. M.*, et al.* 2010. Maximizing productivity of CHO cell-based fed-batch culture using chemically defined media conditions and typical manufacturing equipment. *Biotechnology Progress,* 26**,** 1400.

Jadhav, V.*, et al.* 2013. CHO microRNA engineering is growing up: Recent successes and future challenges. *Biotechnology advances,* 31**,** 1501.

Jayapal, K. P.*, et al.* 2007. Recombinant protein therapeutics from CHO cells-20 years and counting. *Chemical Engineering Progress,* 103**,** 40.

Jeon, M. K.*, et al.* 2011. Combinatorial engineering of ldh-a and bcl-2 for reducing lactate production and improving cell growth in dihydrofolate reductase-deficient Chinese hamster ovary cells. *Applied Microbiology and Biotechnology,* 92**,** 779.

Junker, B.*, et al.* 2015. Quality-by-Design: As Related to Analytical Concepts, Control and Qualification. *Vaccine Analysis: Strategies, Principles, and Control.* Springer.

Kamruzzaman, M.*, et al.* 2012. Non-destructive prediction and visualization of chemical composition in lamb meat using NIR hyperspectral imaging and multivariate regression. *Innovative Food Science & Emerging Technologies,* 16**,** 218.

Kaufman, R. J.*, et al.* 1985. Coamplification and Coexpression of Human Tissue-Type Plasminogen Activator and Murine Dihydrofolate Reductase Sequences in Chinese Hamster Ovary Cells. *Molecular and Cellular Biology,* 5

Khaleghi, B.*, et al.* 2013. Multisensor data fusion: A review of the state-of-the-art. *Information Fusion,* 14**,** 28.

Kneipp, K. & Kneipp, H. 2006. Single molecule Raman scattering. *Focal Point Reviews,* 60.

Kourti, T. & Davis, B. 2012. The business benefits of quality by design (QbD). *Pharm Eng,* 32**,** 52.

Kumar, V. P. & Gupta, N. V. 2015. A Review on quality by design approach (QBD) for Pharmaceuticals. *International Journal of Drug Development and Research,* 7**,** 52.

Landgrebe, D.*, et al.* 2010. On-line infrared spectroscopy for bioprocess monitoring. *Applied Microbiology and Biotechnology,* 88**,** 11.

Laxalde, J.*, et al.* 2014. Combining near and mid infrared spectroscopy for heavy oil characterisation. *Fuel,* 133**,** 310.

Lee, H. W.*, et al.* 2012a. Reproducibility, complementary measure of predictability for robustness improvement of multivariate calibration models via variable selections. *Analytica chimica acta,* 757**,** 11.

Lee, H. W.*, et al.* 2012b. Data fusion-based assessment of raw materials in mammalian cell culture. *Biotechnology and bioengineering,* 109**,** 2819.

Lee, T.-Y*., et al.* 2015. Inhibitory Effect of Excessive Glucose on Its Biochemical Pathway and the Growth of Chinese Hamster Ovary (CHO) Cells. *Journal of Carbohydrate Chemistry,* 34**,** 1.

Ling, W. L. W*., et al.* 2015. Development and manufacturability assessment of chemically-defined medium for the production of protein therapeutics in CHO cells. *Biotechnology progress,* 31**,** 1163.

Liu, J. K. H. 2014. The history of monoclonal antibody development–Progress, remaining challenges and future innovations. *Annals of Medicine and Surgery,* 3**,** 113.

Lourenco, N. D*., et al.* 2012. Bioreactor monitoring with spectroscopy and chemometrics: a review. *Analytical and Bioanalytical Chemistry,* 404**,** 1211.

Lu, F*., et al.* 2013. Automated dynamic fed-batch process and media optimization for high productivity cell culture process development. *Biotechnology and bioengineering,* 110**,** 191.

Lu, X*., et al.* 2011. Application of Mid-infrared and Raman Spectroscopy to the Study of Bacteria. *Food and Bioprocess Technology,* 4**,** 919.

Maleki, M*., et al.* 2007. Multiplicative scatter correction during on-line measurement with near infrared spectroscopy. *Biosystems Engineering,* 96**,** 427.

Martens, H*., et al.* Multivariate linearity transformation for near-infrared reflectance spectrometry. Proceedings of the Nordic symposium on applied statistics, 1983. Stokkand Forlag Publishers Stavanger, Norway, 205.

Matthews, T. E*., et al.* 2016. Closed loop control of lactate concentration in mammalian cell culture by Raman spectroscopy leads to improved cell density, viability, and biopharmaceutical protein production. *Biotechnology and bioengineering,* 113**,** 2416.

Mcfearin, C. L.*, et al.* 2011. Application of fiber-optic attenuated total reflection-FT-IR methods for in situ characterization of protein delivery systems in real time. *Analytical Chemistry,* 83**,** 3943.

Mehdizadeh, H.*, et al.* 2015. Generic Raman-based calibration models enabling real-time monitoring of cell culture bioreactors. *Biotechnology progress,* 31**,** 1004.

Mercier, S. M.*, et al.* 2013. Multivariate data analysis as a PAT tool for early bioprocess development data. *Journal of biotechnology,* 167**,** 262.

Mercier, S. M.*, et al.* 2016. Process analytical technology tools for perfusion cell culture. *Engineering in Life Sciences,* 16**,** 25.

Miller, J. N. & Miller, J. C. 2000. *Statistics and chemometrics for analytical chemistry*.

Miller, K. M.*, et al.* 1996. Quality assurance for biopharmaceuticals - An overview of regulations, methods and problems. *Pharmaceutics Acta Helvetiae,* 71**,** 421.

Milligan, M.*, et al.* 2014. Semisynthetic model calibration for monitoring glucose in mammalian cell culture with in situ near infrared spectroscopy. *Biotechnology and bioengineering,* 111**,** 896.

Mou, Y.*, et al.* 2014. Regularized multivariate scatter correction. *Chemometrics and Intelligent Laboratory Systems,* 132**,** 168.

Murphy, T.*, et al.* Pharmaceutical manufacturing and the quality by design (QBD), process analytical technology (PAT) approach. Signals and Systems Conference (ISSC), 2016 27th Irish, 2016. IEEE, 1.

Nyquist, R. A. & Kagel, R. O. 2012. *Handbook of infrared and raman spectra of inorganic compounds and organic salts: infrared spectra of inorganic compounds*, Academic press.

Olsson, L.*, et al.* 1998. On-line bioprocess monitoring - an academic discipline or an industrial tool? *TrAC Trends in Analytical Chemistry,* 17**,** 88.

Ordás, I.*, et al.* 2012. Anti-TNF monoclonal antibodies in inflammatory bowel disease: pharmacokinetics-based dosing paradigms. *Clinical Pharmacology & Therapeutics,* 91**,** 635.

Pais, D. A.*, et al.* 2014. Towards real-time monitoring of therapeutic protein quality in mammalian cell processes. *Current opinion in biotechnology,* 30**,** 161.

Paudel, A.*, et al.* 2015. Raman spectroscopy in pharmaceutical product design. *Advanced Drug Delivery Reviews,* 89**,** 3.

Rathore, A. S. 2016. Quality by Design (QbD)-Based Process Development for Purification of a Biotherapeutic. *Trends in biotechnology,* 34**,** 358.

Rathore, A. S.*, et al.* 2011. Chemometrics applications in biotech processes: a review. *Biotechnology Progress,* 27**,** 307.

Rathore, A. S.*, et al.* 2015. Fermentanomics: relating quality attributes of a monoclonal antibody to cell culture process variables and raw materials using multivariate data analysis. *Biotechnology Progress,* 31**,** 1586.

Rathore, A. S.*, et al.* 2014. Chemometrics application in biotech processes: assessing comparability across processes and scales. *Journal of Chemical Technology & Biotechnology,* 89**,** 1311.

Rathore, A. S. & Singh, S. K. 2015. Use of Multivariate Data Analysis in Bioprocessing. *Biopharm International,* 28**,** 26.

Rathore, A. S. & Winkle, H. 2009. Quality by design for biopharmaceuticals. *Nature Biotechnology,* 27.

Razinkov, V. I.*, et al.* 2015. Accelerated Formulation Development of Monoclonal Antibodies (mAbs) and mAb-Based Modalities Review of Methods and Tools. *Journal of biomolecular screening,* 20**,** 468.

Read, E. K.*, et al.* 2010. Process Analytical Technology (PAT) for Biopharmaceutical Products: Part I. Concepts and Applications. *Biotechnology and Bioengineering,* 105**,** 276.

Reich, G. 2016. Mid and Near Infrared Spectroscopy. *Analytical Techniques in the Pharmaceutical Sciences.* Springer.

Reichert, J. M. Marketed therapeutic antibodies compendium. MAbs, 2012. Taylor & Francis, 413.

Reichert, J. M. Antibodies to watch in 2014. MAbs, 2014. Taylor & Francis, 5.

Reinhart, D.*, et al.* 2015a. Benchmarking of commercially available CHO cell culture media for antibody production. *Applied Microbiology and Biotechnology,* 99**,** 4645.

Reinhart, D.*, et al.* Influence of cell culture media and feed supplements on cell metabolism and quality of IgG produced in CHO-K1, CHO-S, and CHO-DG44. BMC Proceedings, 2015b. BioMed Central Ltd, P36.

Rhiel, M.*, et al.* 2002. Real-time in situ monitoring of freely suspended and immobilized cell cultures based on mid-infrared spectroscopic measurements. *Biotechnology and bioengineering,* 77**,** 174.

Rinnan, Å.*, et al.* 2009. Review of the most common pre-processing techniques for near-infrared spectra. *Trends in Analytical Chemistry,* 28**,** 1201.

Rodrigues, M. E.*, et al.* 2013. Advances and drawbacks of the adaptation to serum-free culture of CHO-K1 cells for monoclonal antibody production. *Applied biochemistry and biotechnology,* 169**,** 1279.

Rosenberg, M. Transcending the Limits of Manufacturing Medicines. IBC: Production and Economics of Biopharmaceuticals, November 13–15 2000 La Jolla, CA.

Rouiller, Y*., et al.* 2012. Application of Quality by Design to the characterization of the cell culture process of an Fc-Fusion protein. *European Journal of Pharmaceutics and Biopharmaceutics,* 81**,** 426.

Roychoudhury, P*., et al.* 2006a. At-line monitoring of ammonium, glucose, methyl oleate and biomass in a complex antibiotic fermentation process using attenuated total reflectance-mid-infrared (ATR-MIR) spectroscopy. *Analytica Chimica Acta,* 561**,** 218.

Roychoudhury, P*., et al.* 2006b. The potential of mid infrared spectroscopy (MIRS) for real time bioprocess monitoring. *Analytica Chimica Acta,* 571**,** 159.

Roychoudhury, P*., et al.* 2007. Multiplexing fibre optic near infrared (NIR) spectroscopy as an emerging technology to monitor industrial bioprocesses. *Analytica Chimica Acta,* 590**,** 110.

Sales, K. C*., et al.* 2015. In Situ Near-Infrared (NIR) versus High-Throughput Mid-Infrared (MIR) Spectroscopy to Monitor Biopharmaceutical Production. *Applied spectroscopy,* 69**,** 760.

Sandor, M*., et al.* 2013. Comparative study of non-invasive monitoring via infrared spectroscopy for mammalian cell cultivations. *Journal of biotechnology,* 168**,** 636.

Scarff, M*., et al.* 2006. Near Infrared Spectroscopy for bioprocess monitoring and control: Current status and future trends. *Critical Reviews in Biotechnology,* 26**,** 17.

Schenk, J*., et al.* 2007. A simple method to monitor and control methanol feeding of Pichia pastoris fermentations using mid-IR spectroscopy. *Journal of Biotechnology,* 128**,** 344.

Sileoni, V*., et al.* 2013. Evaluation of different validation strategies and long term effects in NIR calibration models. *Food chemistry,* 141**,** 2639.

Silvestri, M*., et al.* 2014. A mid level data fusion strategy for the Varietal Classification of Lambrusco PDO wines. *Chemometrics and Intelligent Laboratory Systems,* 137**,** 181.

Sinelli, N.*, et al.* 2011. Near infrared (NIR) spectroscopy as a tool for monitoring blueberry osmo–air dehydration process. *Food Research International,* 44**,** 1427.

Singh, G. P.*, et al.* 2015. Raman spectroscopy of complex defined media: biopharmaceutical applications. *Journal of Raman Spectroscopy,* 46**,** 545.

Skolnick, P. 2016. Introduction: Biologics to Treat Substance Use Disorders: Vaccines, Monoclonal Antibodies, and Enzymes. *Biologics to Treat Substance Use Disorders.* Springer.

Smith, E. & Dent, G. 2013. *Modern Raman spectroscopy: a practical approach*, John Wiley & Sons.

Sokolov, M.*, et al.* 2015. Fingerprint detection and process prediction by multivariate analysis of fed-batch monoclonal antibody cell culture data. *Biotechnology Progress,* 31**,** 1633.

Sparks, J. R.*, et al.* 2011. Zinc selenide optical fibers. *Advanced Materials,* 23**,** 1647.

Svendsen, C.*, et al.* 2015. Monitoring fermentation processes using in-process measurements of different orders. *Journal of Chemical Technology and Biotechnology,* 90**,** 244.

Swarbrick, B. 2007. Process analytical technology: A strategy for keeping manufacturing viable in Australia. *Vibrational Spectroscopy,* 44**,** 171.

Syvilay, D.*, et al.* 2015. Evaluation of the standard normal variate method for Laser-Induced Breakdown Spectroscopy data treatment applied to the discrimination of painting layers. *Spectrochimica Acta Part B: Atomic Spectroscopy,* 114**,** 38.

Tamburini, E.*, et al.* 2014. Monitoring key parameters in bioprocesses using near-infrared technology. *Sensors,* 14**,** 18941.

Treacy, G. & Knight, D. M. 2016. Monoclonal Antibodies. *Encyclopedia of Immunotoxicology***,** 629.

Trygg, J. & Wold, S. 1998. PLS regression on wavelet compressed NIR spectra. *Chemometrics and Intelligent Laboratory Systems,* 42**,** 209.

Tsao, Y.-S*., et al.* 2005. Monitoring Chinese hamster ovary cell culture by the analysis of glucose and lactate metabolism. *Journal of biotechnology,* 118**,** 316.

Ündey, C*., et al.* 2010. Applied advanced process analytics in biopharmaceutical manufacturing: Challenges and prospects in real-time monitoring and control. *Journal of Process Control,* 20**,** 1009.

Vaidyanathan, S*., et al.* 1999. Monitoring of submerged bioprocesses. *Critical reviews in biotechnology,* 19**,** 277.

Vanarase, A. U*., et al.* 2010. Real-time monitoring of drug concentration in a continuous powder mixing process using NIR spectroscopy. *Chemical Engineering Science,* 65**,** 5728.

Varmuza, K. & Filzmoser, P. 2016. *Introduction to multivariate statistical analysis in chemometrics*, CRC press.

Veith, N*., et al.* 2016. Mechanisms underlying epigenetic and transcriptional heterogeneity in Chinese hamster ovary (CHO) cell lines. *BMC biotechnology,* 16**,** 1.

Vojinovic, V*., et al.* 2006. Real-time bioprocess monitoring Part I: In situ sensors. *Sensors and Actuators B-Chemical,* 114**,** 1083.

Walsh, G. 2010. Biopharmaceutical benchmarks 2010. *Nature Biotechnology* 28.

Whelan, J*., et al.* 2012. In situ Raman spectroscopy for simultaneous monitoring of multiple process parameters in mammalian cell culture bioreactors. *Biotechnology progress,* 28**,** 1355.

Wilkens, C. A*., et al.* 2011. Comparative metabolic analysis of lactate for CHO cells in glucose and galactose. *Biotechnology and Bioprocess Engineering,* 16**,** 714.

Williams, D. H. & Fleming, I. 1995. *Spectroscopic methods in organic chemistry,* London, McGraw-Hill.

Wise, B. M. & Kowalski, B. R. 1995. *Process Analytical Chemistry*.

Wold, S*., et al.* 2001. PLS-regression: a basic tool of chemometrics. *Chemometrics and intelligent laboratory systems,* 58**,** 109.

Wong, N. S*., et al.* 2006. Enhancing recombinant glycoprotein sialylation through CMP-sialic acid transporter over expression in Chinese hamster ovary cells. *Biotechnology and Bioengineering,* 93**,** 1005.

Workman, J*., et al.* 2003. Process analytical chemistry. *Analytical Chemistry,* 75**,** 2859.

Workman, J. J. & Weyer, L. 2012. *Practical Guide and Spectral Atlas for Interpretive Near-Infrared Spectroscopy, Second Edition*, CRC Press.

Yang, C. Enhancing site-specific CHO produced antibody through media optimization using metabolomics approach. *In:* KISS, R., ed. ECI Symposium Series (Cell Culture Engineering XV), 2016.

Yu, L. X. 2008. Pharmaceutical quality by design: product and process development, understanding, and control. *Pharmaceutical Research,* 25**,** 781.

Zagari, F*., et al.* 2013. Lactate metabolism shift in CHO cell culture: the role of mitochondrial oxidative activity. *New biotechnology,* 30**,** 238.

Zhang, L*., et al.* 2005. Multivariate data analysis for Raman imaging of a model pharmaceutical tablet. *Analytica Chimica Acta,* 545**,** 262.

Zhao, L*., et al.* 2015. Advances in process monitoring tools for cell culture bioprocesses. *Engineering in Life Sciences,* 15**,** 459.

Appendix

Table A.1: Process data from CHO 42 batch culture in CD CHO medium [corresponds to Figure 4.1].

Time (Days)	Live Cell Density (Cells per mL)	Live Cell Density Standard Deviation	Cell Viability (%)	Cell Viability Standard Deviation	Glucose (g/L)	Glucose Standard Deviation	Lactate (g/L)	Lactate Standard Deviation
0	251667	6.20	99	0.65	5.93	0.05	0.12	0
1	417778	7.84	98	0.56	5.63	0.12	0.25	0
2	741667	7.25	98	1.05	5.38	0.10	0.43	0.01
3	1108333	3.68	97	1.54	4.98	0.03	0.80	0.01
4	2252778	5.03	97	0.63	3.94	0.10	1.48	0.03
5	3230556	5.65	98	1.32	2.81	0.05	2.14	0.03
6	4219445	4.44	97	1.11	1.97	0.03	2.33	0.01
7	5827778	9.14	95	0.87	1.15	0.00	2.58	0.02
8	5233334	4.31	91	2.22	0.66	0.01	2.69	0.01
9	3644445	7.12	69	3.73	0.42	0.01	2.63	0.03
10	2855556	6.90	56	4.25	0.27	0.00	2.51	0.03
11	1555556	6.02	35	2.91	0.15	0.00	2.54	0.06
12	991667	5.63	27	4.87	0.12	0.00	2.46	0.03
13	0	1.00	2	3.00	0.15	0.00	2.73	0.05

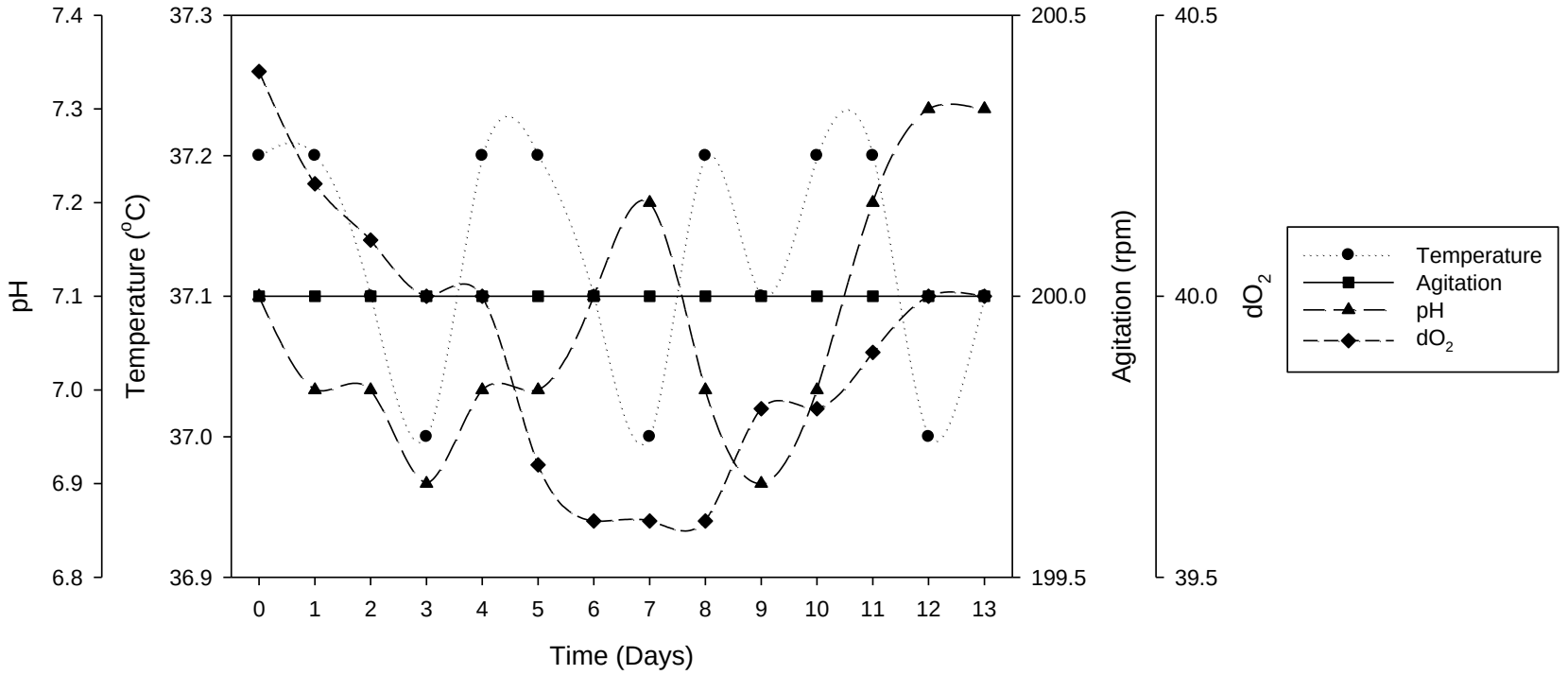


Figure A.1: Bioreactor process data from CHO 42 batch culture in CD CHO medium.

Table A.2: Process data from CHO 42 fed batch culture in CD CHO medium with EfficientFeed™ B [corresponds to Figure 4.17].

Time (Days)	Live Cell Density (Cells per mL)	Live Cell Density Standard Deviation	Cell Viability (%)	Cell Viability Standard Deviation	Glucose (g/L)	Glucose Standard Deviation	Lactate (g/L)	Lactate Standard Deviation
0	243333	4.15	98	1.14	5.30	0.05	0.10	0
1	489167	3.94	98	1.61	5.80	0.06	0.24	0
2	881250	2.29	98	0.49	5.55	0.11	0.45	0.01
3	1666667	3.46	98	1.16	5.06	0.02	0.84	0.00
4	2983333	9.98	98	0.70	5.27	0.09	1.43	0.03
5	4858334	10.69	97	1.08	4.05	0.03	2.21	0.02
6	5008334	8.95	95	1.72	3.65	0.03	3.09	0.03
7	5241667	6.03	94	1.45	2.06	0.02	3.93	0.04
8	4700000	5.10	87	2.79	2.24	0.01	4.06	0.02
9	3183333	5.06	47	1.15	1.70	0.03	4.30	0.03
10	2150000	1.80	40	1.83	2.66	0.04	4.58	0.04
11	1741667	9.02	37	5.98	2.95	0.01	3.89	0.01
12	608333	6.81	25	7.21	3.98	0.02	3.62	0.03
13	0	0.00	0	0.00	3.56	0.03	3.48	0.03

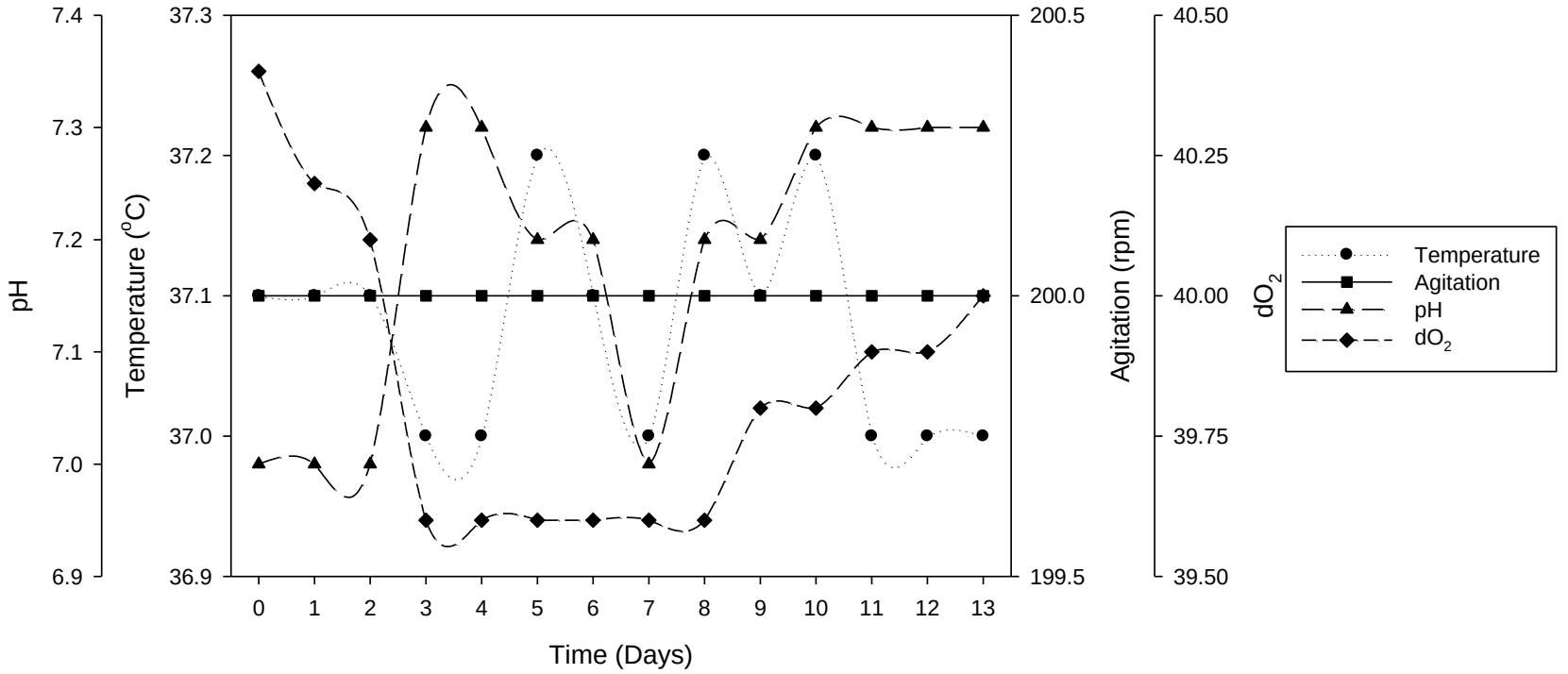


Figure A.2: Bioreactor process data from CHO 42 fed batch culture in CD CHO medium with EfficientFeed™ B.

Table A.3: Process data from CHO 42 batch culture in CD OptiCHO medium [corresponds to Figure 5.1].

Time (Days)	Live Cell Density (Cells per mL)	Live Cell Density Standard Deviation	Cell Viability (%)	Cell Viability Standard Deviation	Glucose (g/L)	Glucose Standard Deviation	Lactate (g/L)	Lactate Standard Deviation
0	223333	4.07	99	1.61	4.59	0.03	0.06	0
1	388333	6.22	97	1.41	5.08	0.09	0.21	0
2	552500	4.77	96	0.86	4.93	0.01	0.33	0.00
3	795833	7.99	92	2.38	4.61	0.03	0.56	0.00
4	1257500	5.95	83	3.13	3.36	0.01	0.76	0.35
5	2328333	2.93	88	0.74	2.91	0.02	1.49	0.02
6	2477083	8.89	91	1.67	1.77	0.01	2.07	0.04
7	2433333	3.71	82	2.98	0.93	0.01	2.36	0.02
8	1787500	3.79	76	2.94	0.32	0.01	2.55	0.03
9	1725000	2.52	60	2.13	0.02	0.00	2.62	0.04
10	991667	3.86	37	2.46	0.02	0.01	2.51	0.03
11	766667	4.25	26	2.93	0.02	0.00	2.47	0.12
12	362500	2.87	14	1.82	0.02	0.00	2.29	0.04
13	145833	1.15	6	1.06	0.02	0.00	2.27	0.06

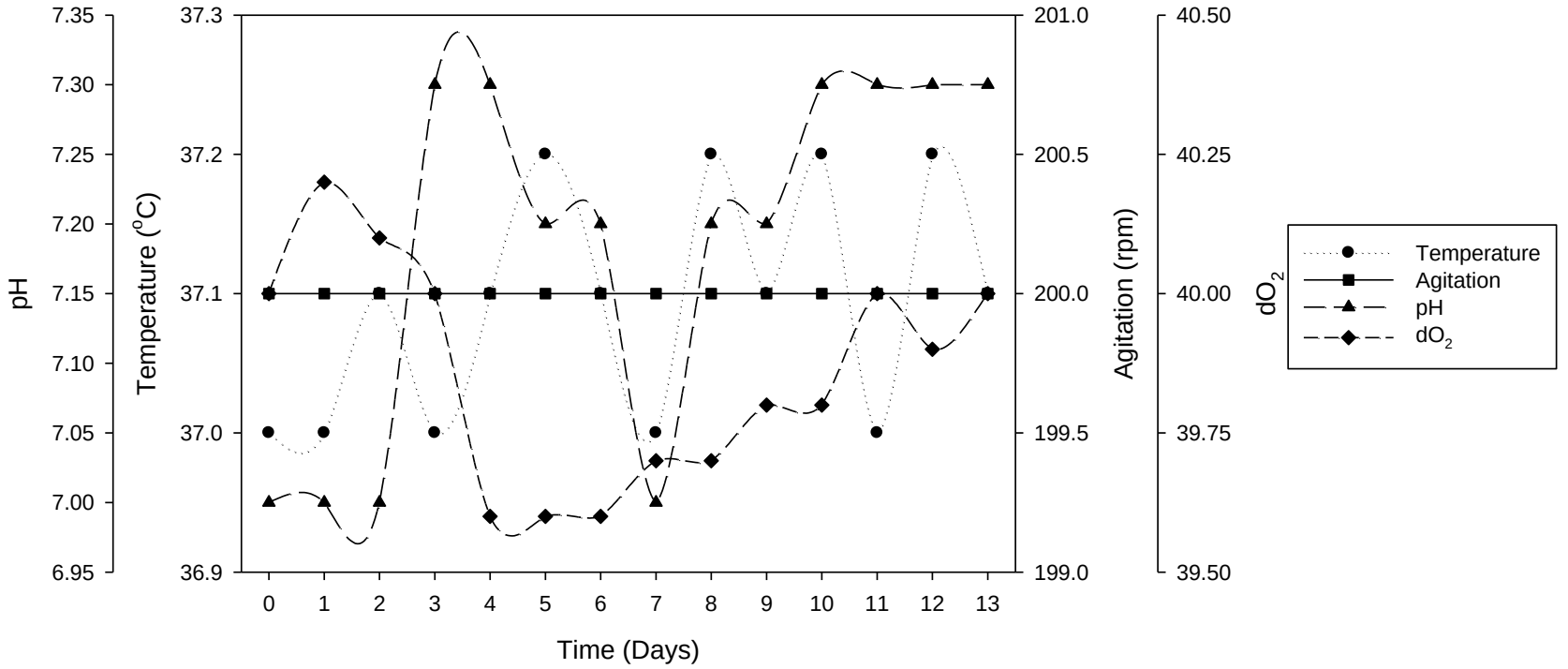


Figure A.3: Bioreactor process data from CHO 42 batch culture in CD OptiCHO medium.

Table A.4: Process data from CHO 42 fed batch culture in CD OptiCHO medium with EfficientFeed™ A [corresponds to Figure 5.12].

Time (Days)	Live Cell Density (Cells per mL)	Live Cell Density Standard Deviation	Cell Viability (%)	Cell Viability Standard Deviation	Glucose (g/L)	Glucose Standard Deviation	Lactate (g/L)	Lactate Standard Deviation
0	335000	3.05	95	0.75	5.23	0.17	0.10	0
1	508333	5.24	96	1.30	4.90	0.13	0.28	0
2	1008333	5.97	97	1.10	4.84	0.05	0.46	0.01
3	1318750	4.27	95	2.67	4.29	0.02	0.75	0.01
4	2616667	10.25	96	1.01	5.46	0.03	1.50	0.02
5	2558333	7.54	95	2.01	2.36	0.03	2.31	0.03
6	2683333	2.39	74	2.44	2.86	0.02	2.70	0.01
7	16667	0.29	1	0.23	2.85	0.03	1.92	0.02

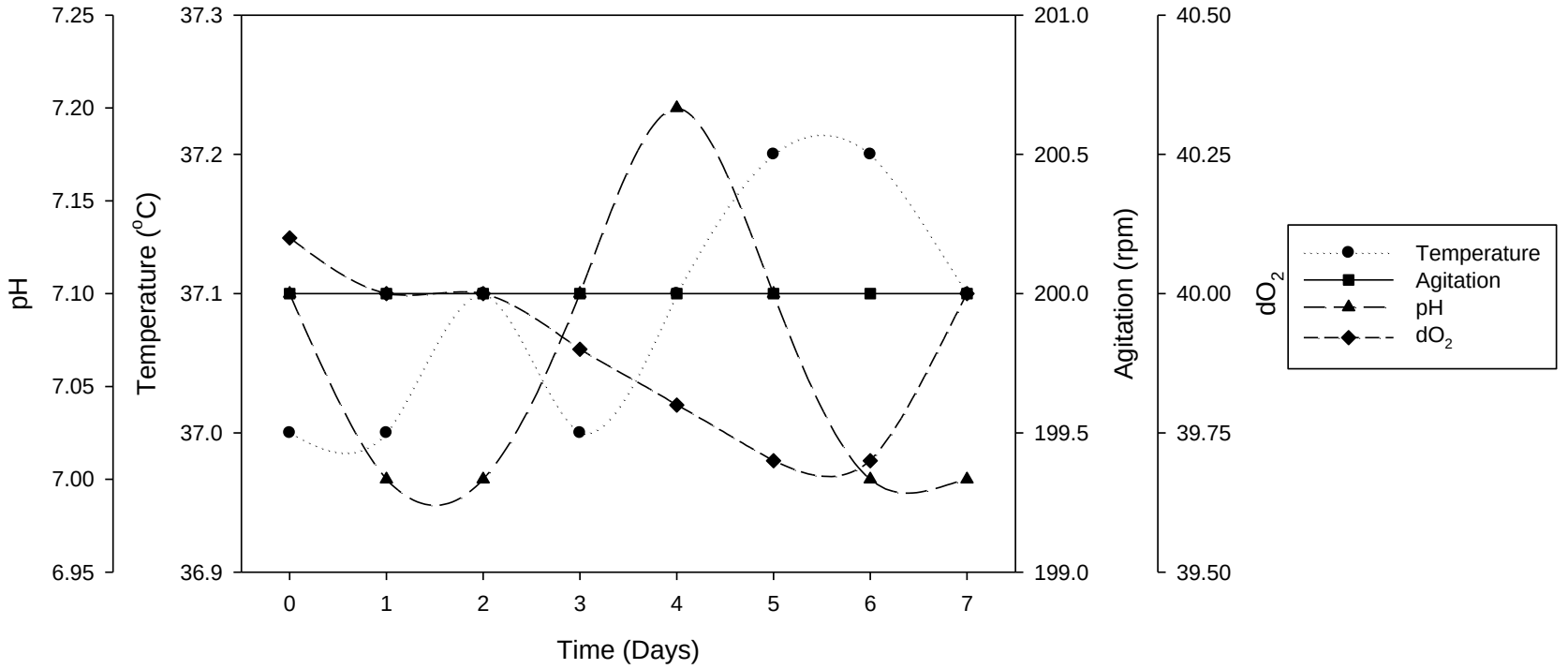


Figure A.4: Bioreactor process data from CHO 42 fed batch culture in CD OptiCHO medium with EfficientFeed™ A.

Table A.5: Process data from CHO 42 batch culture in Dynamis medium [corresponds to Figure 6.1].

Time (Days)	Live Cell Density (Cells per mL)	Live Cell Density Standard Deviation	Cell Viability (%)	Cell Viability Standard Deviation	Glucose (g/L)	Glucose Standard Deviation	Lactate (g/L)	Lactate Standard Deviation
0	257500	1.80	100	0.59	5.92	0.11	0.14	0
1	469167	7.39	97	1.82	5.90	0.11	0.27	0
2	791667	4.41	96	0.56	5.63	0.03	0.42	0
3	939583	5.43	95	1.76	5.24	0.04	0.71	0.01
4	1289583	4.34	94	0.81	4.46	0.04	1.13	0.01
5	1379167	5.14	88	2.17	3.41	0.03	1.70	0.01
6	1741667	4.08	87	0.57	2.31	0.02	2.33	0.01
7	1975000	6.43	83	2.07	1.53	0.01	2.60	0.02
8	1641667	8.00	66	2.43	1.12	0.01	2.61	0.03
9	1341667	3.59	56	2.30	0.68	0.00	2.72	0.04
10	1225000	1.05	39	1.44	0.24	0.00	2.88	0.04
11	933333	1.26	34	0.06	0.03	0.00	2.86	0.03
12	1283333	1.53	53	1.68	0.02	0.01	2.58	0.09
13	1008333	1.53	41	5.02	0.02	0.00	2.41	0.03
14	758333	3.06	27	1.09	0.02	0.01	2.33	0.03

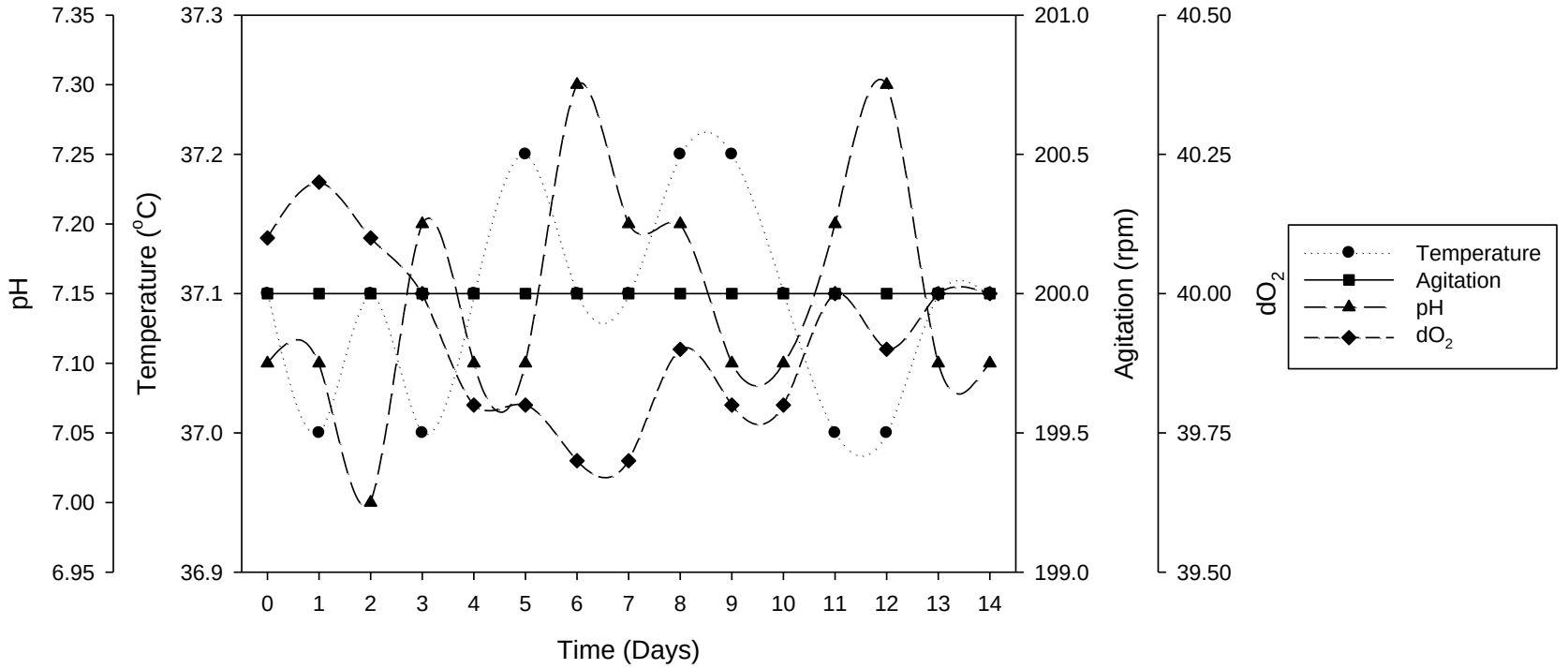


Figure A.5: Bioreactor process data from CHO 42 batch culture in Dynamis medium.