

The Use of Infrared Spectroscopy to Monitor Bio-catalytic Processes

Peter Gardner

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

September 2012

Copyright Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Signed:

Date:

Acknowledgements

First and foremost, I would like to thank my supervisors: Professor Brian McNeil, Dr. Linda Harvey & Dr. S. Alison Arnold for their patience, knowledge, encouragement and guidance over the last four years. Also I would like to express my gratitude to Dr. Alison Nordon for all her help and patience with the modelling aspects of the project and the Matlab software package.

I would like to thank Ingenza Ltd for supplying all the organisms and enzymes throughout the project. Thanks also to the staff at Ingenza Ltd., specifically Dr. Reuben Carr, Dr. Fraser Brown and Dr. Ian Archer, for all their assistance in carrying out the biotransformation processes and with the associated analytical methods.

Thank you also to my friends in the Strathclyde Fermentation Centre: Dr. Manal Eshelli, Dr. Ebtihaj Jambi, Dr. Tantima Kumlung, Ivo Kretzers, Melissa Black, Brian Griffin, Laura Jeffrey and also Walter McEwan for the technical support throughout. I would like to express specific gratitude to Dr. Ioannis Vouglaris and Dr. Mariana Fazenda for their patience, and persistence, in teaching me how to carry out the various laboratory techniques that allowed this project to happen.

A special thanks also to my parents, family and friends for their continued support.

I am grateful to Fibre Photonics for the loan of the mid infrared spectrometer system and to the department of pure and applied chemistry at Strathclyde for a loan of the associated ATR probe. This project was funded by a BBSRC industrial CASE award.

Table of Contents

Copyright Declaration.....	i
Acknowledgements.....	ii
Table of Contents.....	iii
List of Figures.....	xii
List of Tables.....	xxiv
Abbreviations.....	xxvii
Abstract.....	xxix
1. Literature Review.....	1
1.1 Introduction.....	1
1.1.1 Process Analytical Technology Initiative.....	2
1.1.2 Bioprocesses.....	3
1.1.2.1 Chemical Synthesis vs. Biocatalysis.....	4
1.1.2.1.1 Process Considerations.....	4
1.1.2.1.2 Bioreactor.....	5
1.1.2.1.3 Enzyme Form.....	5
1.1.2.2 Industrial Applications.....	5
1.1.3 Biotransformation Process.....	6
1.2 Spectroscopic Approaches to Process Monitoring.....	6
1.2.1 Near Infrared Spectroscopy.....	9
1.2.1.1 Spectral Acquisition.....	10
1.2.1.2 Near Infrared Instrumentation.....	11
1.2.1.3 Off-line/At-line Sampling.....	12
1.2.1.4 In-line Sampling.....	12
1.2.1.5 In-situ Sampling.....	14

1.2.2 Middle Infrared Spectroscopy	16
1.2.2.1 Off-line/At-line Sampling.....	18
1.2.2.2 In-line Sampling	19
1.2.2.3 In-situ Sampling.....	20
1.3 Process Analytical Technology	21
1.3.1 Submerged Culture Bioprocesses	22
1.3.1.1 Conventional Approaches.....	22
1.3.1.2 Infrared Spectroscopy	23
1.3.1.2.1 Analyte Selection.....	23
1.3.1.2.1.1 Monitoring of Substrates & Products	23
1.3.1.2.1.2 Monitoring of Culture Medium	25
1.3.1.2.1.3 Monitoring of Physical Parameters	26
1.3.1.2.2 Spectral Pre-processing.....	26
1.3.1.2.3 Calibration Strategy	27
1.3.1.2.4 Near Infrared.....	29
1.3.1.2.5 In-situ Near Infrared	30
1.3.1.2.5.1 Multiplexing	31
1.3.1.2.6 Mid Infrared.....	33
1.3.2 Biotransformation Processes	34
1.3.2.1 Near Infrared Spectroscopy	34
1.3.2.2 Mid Infrared Spectroscopy	36
1.3.3 Alternative Process Monitoring Approaches.....	39
1.4 Chiral Amino Acids.....	40
1.5 Aims & Objectives	41
2. Background & Theory	43
2.1 D-amino Acid Oxidases.....	43

2.2 Transaminase Enzymes.....	44
2.3 High Performance Liquid Chromatography	45
2.4 Vibrational Spectroscopy.....	46
2.4.1 Middle Infrared.....	51
2.4.2 Near Infrared.....	52
2.4.3 Instrumentation	53
2.4.3.1 Dispersive Instruments	53
2.4.3.2 Fourier Transform Instruments.....	55
2.5 Chemometrics	56
2.5.1 Principal Component Analysis	56
2.5.2 Spectral Pre-processing.....	58
2.5.2.1 Mean Centering	58
2.5.2.2 Smoothing & Derivatives	59
2.5.2.3 Other Pre-processing Methods.....	59
2.5.3 Partial Least Squares.....	60
2.5.4 Artificial Neural Networks	61
2.5.5 Model Evaluation.....	61
2.5.5.1 Internal Validation	62
2.5.5.2 External Validation	62
2.5.6 Calibration Transfer.....	63
3. Materials & Methods	64
3.1 Standard Preparation.....	64
3.1.1 Commercially Available Standards	64
3.1.2 Non-commercially Available Standards.....	65
3.1.2.1 Tri-methyl Pyruvic Acid.....	65
3.1.2.2 Tertiary-butyl Glycine	65

3.1.2.3 Compound B	66
3.2 High Performance Liquid Chromatography	67
3.2.1 Amino Acid Separation & Quantification	68
3.2.1.1 Derivatisation Procedure.....	68
3.2.1.1.1 Preparation of Borate Buffer.....	69
3.2.1.1.2 Preparation of Derivatisation Reagent.....	70
3.2.1.1.3 Derivatisation of Amino Acid	70
3.2.1.2 Chromatographic Separation	71
3.2.1.2.1 Tertiary-butyl Glycine	71
3.2.1.2.2 Other Amino Acids.....	71
3.2.2 Keto Acid Separation & Quantification.....	72
3.2.2.1 Tri-methyl Pyruvic Acid.....	72
3.2.2.2 Other Keto-acids	72
3.2.2.3 Organic Acids	72
3.2.2.4 Method Validation & Error Estimation	73
3.3 Biotransformation Processes	75
3.3.1 Artificial Spiking	75
3.3.2 Tertiary-butyl Glycine De-racemisation.....	76
3.3.2.1 Preparation of Tertiary-butyl Glycine.....	76
3.3.2.2 Biotransformation	78
3.3.3 Amino Butyric Acid De-racemisation	81
3.3.4 Alanine De-racemisation	83
3.3.5 In-situ Biotransformation.....	84
3.3.5.1 Preparation of Cell Bank	84
3.3.5.2 E. coli Fermentation.....	86
3.3.5.2.1 Biomass Estimation	89

3.3.5.3 Transaminase Biotransformation.....	90
3.4 Mid Infrared Measurements.....	92
3.4.1 At-line Mid Infrared.....	92
3.4.2 In-situ Middle Infrared.....	92
3.5 Near Infrared Measurements	93
3.5.1 At-line Near Infrared Measurements	93
3.5.2 In-situ Process Measurements	94
3.6 Spectral Manipulation & Modelling.....	95
4. Amino Acid De-racemisation	96
4.1 Process Overview	96
4.2 Aims & Objectives	97
4.2.1 Novelty	98
4.3 Analytical Reference Method	99
4.3.1 Amino Acid	99
4.3.1.1 Method Development	100
4.3.1.2 Method Validation	102
4.3.2 Keto Acid.....	106
4.3.2.1 Method Development	106
4.3.2.2 Method Validation	107
4.4 Initial Feasibility Study.....	108
4.4.1 Near Infrared Spectroscopy	108
4.4.2 Mid Infrared Spectroscopy	115
4.4.3 Experimental Design.....	120
4.5 Calibration Model.....	122
4.6 Model Evaluation.....	135
4.7 Model Application	143

4.8 Summary & Conclusions	150
5. Small Scale Biotransformation	153
5.1 Process Overview	153
5.2 Aim & Objective.....	154
5.2.1 Novelty	154
5.3 Analytical Reference Method	155
5.3.1 Amino Butyric Acid	155
5.3.1.1 Method Development	155
5.3.1.2 Method Validation	156
5.3.2 Keto Butyric Acid.....	157
5.3.2.1 Method Development	157
5.3.2.2 Method Validation	158
5.4 Initial Feasibility Study.....	159
5.5 Process Replicates	165
5.6 Near Infrared.....	167
5.6.1 Calibration Model & Validation.....	167
5.6.2 Model Contributors.....	172
5.6.2.1 Combination Bands	172
5.6.2.2 First Overtones.....	175
5.6.2.3 Combined Model	179
5.6.2.4 Near Infrared Summary	180
5.7 Mid Infrared Spectroscopy	181
5.7.1 Mid Infrared Raw Spectra.....	181
5.7.2 Mid Infrared First Derivative.....	186
5.7.3 Genetic Algorithm	192
5.7.3.1 Raw Dataset	192

5.7.3.2 First Derivative Dataset	194
5.8 Model Summary	196
5.9 Process Application	199
5.9.1 Middle Infrared	199
5.10 Conclusions.....	203
6. Alanine De-racemisation	204
6.1 Process Overview	204
6.2 Aim & Objective.....	206
6.2.1 Novelty	206
6.3 Analytical Reference Methods.....	207
6.3.1 Quantification of Alanine	207
6.3.2 Quantification of Organic Acids.....	208
6.4 Initial Feasibility Study.....	209
6.4.1 Pure Components.....	209
6.4.2 Parameter Optimisation	212
6.4.3 Summary.....	213
6.5 Biotransformation Replicates	214
6.5.1 Experimental Design.....	214
6.5.2 Process Replicates	215
6.5.2.1 Initial Observations.....	216
6.5.2.2 Data Interpolation	219
6.5.2.2.1 Linear Interpolation	219
6.5.2.2.2 Exponential Decay.....	220
6.5.2.2.3 Kinetic Interpolation.....	221
6.5.2.3 Alanine Model	223
6.5.2.1 Organic Acid Model	232

6.6 Conclusions.....	235
7. Fermentation for <i>In-situ</i> Biotransformation.....	236
7.1 Process Summary.....	236
7.1.1 Aim & Objective.....	237
7.1.2 Novelty	237
7.2 Biomass Model.....	238
7.3 Fingerprinting & Control.....	241
7.3.1 Initial Observations.....	241
7.3.2 Pattern Recognition Techniques	243
7.3.3 Standard Isolinear Method of Class Assignment.....	244
7.3.4 Partial Least Squares – Discriminant Analysis.....	246
7.3.5 Artificial Neural Network.....	248
7.3.6 Summary.....	251
7.4 Conclusions.....	253
8. <i>In-situ</i> Biotransformation Process	255
8.1 Process Overview	255
8.2 Aim & Objective.....	257
8.2.1 Novelty	257
8.3 Analytical Reference Method	258
8.3.1 Preparation of Standards.....	258
8.3.2 Amino Acid Quantification	261
8.3.2.1 Method Validation	262
8.3.3 Organic Acid Quantification.....	263
8.3.3.1 Method Validation	263
8.4 Initial Feasibility.....	264
8.4.1 In-situ Near Infrared	264

8.4.2 At-line Middle Infrared.....	267
8.5 The Biotransformation Process.....	269
8.5.1 Middle Infrared Model.....	269
8.5.2 Model Refinement	275
8.5.3 Calibration Transfer.....	281
8.5.3.1 Sample Selection & Identification.....	282
8.5.3.2 Direct Standardisation.....	286
8.5.3.3 Piecewise Direct Standardisation.....	288
8.5.3.4 Spectral Space Transformation.....	292
8.5.4 Model Summary	296
9. Conclusions.....	300
10. Future Work.....	304
11. References.....	307
Appendix I	317
Appendix II.....	318
11.1 TMP Validation Data.....	318
11.2 ABA Validation Data	321
11.3 Keto Butyric Acid.....	324
11.4 L-alanine & Compound B Validation Data.....	327
11.5 Compound A.....	330
11.6 Pyruvic Acid & Acetic Acid.....	333
Appendix III.....	336

List of Figures

Figure 1-1 – Graphical representation of the various sampling approaches (Vaidyanathan <i>et al.</i> (1999)) utilised for the acquisition of the infrared spectra of the reactor contents at various stages during a bioprocess. ¹⁹	7
Figure 1-2 – Illustration of the three possible infrared measurement approaches: (a) transmission, (b) reflectance and (c) transreflectance. Adapted from Vaidyanathan <i>et al.</i> (1998). ²⁰	10
Figure 1-3 - Schematic representation adapted from Holm-Nielsen <i>et al.</i> (2007) of typical in-line flow through system employed for sampling and spectral acquisition in the near infrared region. ³⁹	13
Figure 1-4 - Schematic representation and image of transreflectance probe for use with near infrared spectroscopy adapted from Roychoudhury <i>et al.</i> (2007), Vaidyanathan <i>et al.</i> (1998) and Hassell <i>et al.</i> (1998). ^{20, 42, 46} Key components of the probe have also been indicated.	15
Figure 1-5 – Illustration of refraction and total internal reflectance of light at the boundary of materials and propagation of light through an ATR crystal illustrating the evanescent wave passing into the sample. In this case n_1 refers to the refractive index of the crystal material, n_2 the refractive index of the sample material, θ_i the angle of incidence and θ_r the angle of refraction. ^{58, 59}	18
Figure 1-6 - Proposed mechanism for the conversion from α -glucose to β -glucose occurring in solution. ⁷⁸	24
Figure 1-7 – Schematic representation of multiplexing setup adapted from Chen <i>et al.</i> (2011). ⁸⁷ One spectrometer system is used to acquire the infrared spectra from three different reactor systems. Spectra can then be used with previously constructed multivariate models for the quantification of multiple analytes in each reactor.	31
Figure 1-8 - Mechanism of typical Baeyer-Villiger reaction carried out chemically.	35
Figure 1-9 – Conversion of ketone to lactone using microbial Baeyer-Villiger mono-oxidase enzyme as described by Doig <i>et al.</i> (2003) and monitored spectroscopically by Bird <i>et al.</i> (2002). ^{92, 93}	35

Figure 1-10 - Routes for the degradation of cyanide containing compounds to the corresponding amine and carboxylic acid by <i>Rhodococcus rhodochrous</i> reported by Dadd <i>et al.</i> (2000). ⁹⁵	36
Figure 1-11 – Biotransformation process for the conversion of 1-phenylpropane-1,2-dione to (1R,2S)-1-phenylpropane-1,2-diol used in the synthesis of the stimulant ephedrine. ⁷²	37
Figure 2-1 – Proposed mechanism for the conversion of D-amino acid substrate into the corresponding α -keto acid proceeding via an α -imino acid intermediate and catalysed by DAAO enzyme and FAD adapted from Garcia-Garcia <i>et al.</i> (2008) ¹⁰⁵	44
Figure 2-2 - Transaminase biotransformation process illustrating conversion of substrate keto acid to the desired amino acid with retention of stereochemistry.	45
Figure 2-3 - Schematic representation of the main components in a HPLC system adapted from Hamilton <i>et al.</i> (1982) ¹¹⁰	45
Figure 2-4 - Illustration of possible vibrations of bonds within a molecule adapted from Kemp (1987) ¹¹³	48
Figure 2-5- Schematic representation of the electronic, vibrational and rotational energy levels adapted from Crabb <i>et al.</i> (1995). ⁵⁰	48
Figure 2-6 – Plot of potential energy against bond length for a simple diatomic molecule behaving as a harmonic oscillator adapted from Hollas (2002). ¹¹⁶	49
Figure 2-7 – Plot of potential energy against bond length for simple diatomic molecule behaving as an anharmonic oscillator. Energy levels are no longer equally distributed and bond dissociation is taken into consideration.	50
Figure 2-8 - Diagrammatic representation of vibrational transitions on absorption of appropriate light at frequency in the mid infrared region.	51
Figure 2-9 - Overtone transition from the ground energy level to the second excited state.	52
Figure 2-10 - Schematic representation of basic layout of dispersive infrared spectrometer and the component parts of monochromator unit illustrating how incident beam of light travels through the spectrometer system.	54
Figure 2-11 - Schematic representation of the key components of an FT-IR spectrometer and the Michelson interferometer.	55

Figure 2-12 - Diagrammatic representation of data reduction by principal component analysis on a simple two-dimensional dataset.	57
Figure 3-1 - Production of compound B for use as reference standard from the substituted pyrrole parent compound in an adaptation of the procedure described by Clark <i>et al.</i> (2001). ¹²⁵	66
Figure 3-2 – Gilson modular HPLC system fitted with both refractive index and DAD detector systems.	68
Figure 3-3 - Derivatisation of amino acid with <i>ortho</i> -phthalaldehyde (OPA) and 2-Mercaptoethanol to form chromophore containing compound allowing detection by DAD at $\lambda_{\text{max}}=230\text{nm}$ proceeding via mechanism suggested by Wong <i>et al.</i> (1985). ¹²⁹	69
Figure 3-4 - Reductive amination process for the conversion of TMP to racemic TBG using ammonia solution under a pressurised hydrogen atmosphere catalysed by 5% palladium on carbon.....	76
Figure 3-5 – Image of Applikon ADI bioreactor system illustrating the key components of the system.....	79
Figure 3-6 - De-racemisation of D/L-ABA using the freeze-dried DAAO enzyme. 81	
Figure 3-7 - Image of Braun Biostat Q bioreactor system illustrating the key components of the system.....	82
Figure 3-8 - De-racemisation of D/L-alanine to L-alanine and pyruvic acid using DAAO enzyme contained within a <i>Pichia pastoris</i> host cell.	83
Figure 3-9 - Image of BIOSTAT C-DCU bioreactor system illustrating the locations of the key components and features of the bioreactor system.	86
Figure 3-10 - Biotransformation process for the generation of compound B from L-alanine (absolute stereochemistry S-alanine) and compound A using a transaminase enzyme contained within the whole cell E-Coli resulting from the fermentation process (3.3.5.2).....	91
Figure 4-1 - Biotransformation process for the de-racemisation of TBG by selectively converting the D-TBG enantiomer to TMP leaving L-TBG untouched..	96
Figure 4-2 - Sample chromatogram of TBG analysed using the developed HPLC method and expansion of peak for the calculation of peak asymmetry factor and peak tailing factor.....	102

Figure 4-3 - Calibration curve with linear line of best fit through the data obtained for the calibration samples analysed using the developed TBG method. (Equation of straight line $Y=10429358.2x + 7892698$).....	103
Figure 4-4 - Calibration curve for TMP, showing linear line of best fit through the calibration data.....	107
Figure 4-5 - Second derivative spectra of pure component solutions of the amino acid (TBG) and keto acid (TMP) plotted against the second derivative spectrum of distilled water. Regions were expanded to illustrate the (i) second overtones region (ii) first overtones region [dominant water signals have been removed for clarity] and (iii) the combination bands region.	110
Figure 4-6 – First overtones (i) and combination band regions (ii) of near infrared spectra acquired of solutions containing a mixture of amino acid and keto acid at various concentrations (detailed in Table 7) typically expected during the biotransformation process.....	111
Figure 4-7 – Fingerprint region second derivative pure component spectra of the two key analytes of interest and water background matrix compared to identify useful regions for use in construction of a multivariate model for the biotransformation process.	118
Figure 4-8 – Fingerprint mid infrared region of samples containing a mixture of TBG and TMP at varying concentrations in a manner representative of the biotransformation process as it progressed to identify regions that exhibited spectral changes for each analyte with concentration. Arrows denote spectral features that demonstrate an increase or decrease in signal intensity with concentration as detailed in Table 7.	120
Figure 4-9 - Plots of measured concentration (mMol dm^{-3}) (from reference analysis) against predicted concentration (mMol dm^{-3}) (from leave one out cross validation) for both the analytes of interest in the synthetic experimental design model constructed using the selected fingerprint region of the mid infrared spectra.....	122
Figure 4-10 - Plot of scores associated with principal component 2 plotted against scores associated with principal component 1 used to identify the most suitable datasets for calibration and the most suitable for validation of the constructed model.	124

Figure 4-11 – Amended regions removing features identified as being redundant or common to both analytes of interest to improve the model constructed for the system. Regions retained were between 900-1031 cm^{-1} , 1100-1340 cm^{-1} and 1440-1620 cm^{-1}	128
Figure 4-12 – Plots of predicted analyte concentration (mMol dm^{-3}) against measured analyte concentration (mMol dm^{-3}) for the validation biotransformation replicates. (i) and (ii) Analyte concentration predictions for validation datasets based on models constructed using the triplicate spectra and raw HPLC reference data values. (iii) and (iv) Analyte concentration predictions for validation datasets using only the single spectra and the mean reference data value.....	134
Figure 4-13 - Loadings associated with latent variable one in the constructed PLS model. Overlaid are plots of the pure component amino acid (TBG) and keto acid (TMP) for comparison.	136
Figure 4-14 - Loadings associated with latent variable two in the constructed PLS model for the biotransformation process. Again overlaid are the pure component spectra of the amino acid (TBG) and keto acid (TMP) for comparison.	137
Figure 4-15 - Regression coefficients associated with the constructed PLS model plotted alongside the pure component spectra of the amino acid (TBG) and keto acid (TMP) for comparison. (i) Plots the first regression coefficient against the pure component spectra. (ii) Plots the second regression coefficient with the pure component spectra.	141
Figure 4-16 – Predicted concentration (mMol dm^{-3}) plotted against measured concentration (mMol dm^{-3}) of both amino acid and keto acid for biotransformation processes carried out at different site from those used to construct and validate the PLS model.	146
Figure 4-17 - Principal component analysis carried out on samples used in the construction of the calibration model (that were carried out in the Applikon bioreactor system) and those carried out off-site in the heated water jacket vessel.	148
Figure 4-18 - Summary plots of concentrations of TBG and TMP determined from the reference method (HPLC) and the PLS model for (i) validation set 1 and (ii) validation set 2. HPLC values for TBG are quoted as $\pm 0.27 \text{ mMol dm}^{-3}$ and TMP as	

±0.14 mMol dm ⁻³ as calculated. Concentrations predicted from model were quoted as ±25.8 mMol dm ⁻³ and ±11.9 mMol dm ⁻³ respectively.....	151
Figure 5-1 – Schematic representation of the biotransformation of D-amino acid (ABA) to the keto acid (KBA) using the DAAO enzyme contained within a <i>Pichia pastoris</i> host.....	153
Figure 5-2 - Calibration curve for quantification of ABA showing linear line of best fit through the data.....	156
Figure 5-3 – Calibration curve for the quantification of KBA using the developed HPLC method.	158
Figure 5-4 - Second derivative spectrum of the pure component spectra of ABA, KBA and water obtained using the at-line near infrared spectrometer system. (i) Displays the second overtones region, (ii) the first overtones and (iii) the combination bands region.....	162
Figure 5-5 - Raw mid infrared spectrum of pure component solutions of amino acid (ABA), keto acid (KBA), water and a mixture of the analytes.	164
Figure 5-6 - First derivative spectra of pure component samples of amino acid (ABA), keto acid (KBA), the background water matrix and a mixture of the two analytes at concentrations representative of those expected during the biotransformation process.....	164
Figure 5-7 – Plots of measured concentration (mMol dm ⁻³) (from reference analysis method) against predicted concentration (mMol dm ⁻³) (from PLS model) for the two validation datasets using the various spectral regions of interest. Figures (i) and (ii) represent the model constructed from the combination bands region, (iii) and (iv) the first overtones region, (v) and (vi) the model constructed from the combined regions. Points shown in grey are the measured concentrations plotted against predicted from the internal validation procedure	171
Figure 5-8 - (i) Plot of first regression coefficient and pure component spectra of ABA and KBA. (ii) Plot of second regression coefficient and pure component spectra of the key analytes of interest. (iii) Both regression coefficients obtained for the model plotted for comparison. (iv) Plot of first regression coefficient and inverse of second regression coefficient to highlight subtle variations in the regression coefficients, which result from the subtle shifts in the pure component spectra.	173

Figure 5-9 – (i) Plot of first regression coefficient and pure component spectra in first overtones region. (ii) Second regression coefficient along with pure component spectra in the first overtones region. (iii) Both regression coefficients associated with the model plotted together for comparison. (iv) Plot of first regression coefficient and inverse of second regression coefficient to identify any subtle shifts in the regression coefficients.....	177
Figure 5-10 - (i) Plot of first regression coefficient along with second derivative spectrum of pure component ABA and KBA. (ii) Plot of second regression coefficient and pure component ABA and KBA samples. (iii) Regression coefficients associated with the constructed model plotted for comparison. (iv) First regression coefficient and inverse of second regression coefficient overlaid for comparison of peak positions.	178
Figure 5-11 - Plot of RMSEC / RMSECV values against latent variable number used to determine the optimum number of latent variables for the constructed PLS model based on the mid infrared spectra without any spectral pre-processing applied.	182
Figure 5-12 - Plots of measured concentration (mMol dm^{-3}) (from reference analysis method) against predicted concentration (mMol dm^{-3}) (from the PLS model) for both the amino acid (i) and keto acid (ii) obtained from the two external validation datasets.....	183
Figure 5-13 - (i) Regression coefficient 1 plotted alongside the pure component spectrum of ABA and KBA. (ii) Second regression coefficient plotted with pure component ABA and KBA spectra. (iii) Both regression coefficients associated with the constructed model superimposed. (iv) First regression coefficient and inverse of second regression coefficient superimposed.....	185
Figure 5-14 – Measured concentration (mMol dm^{-3}) (from reference analysis method) plotted against predicted concentration (mMol dm^{-3}) (from PLS model) for external validation datasets. (i) Illustrates the amino acid (ABA) component whilst (ii) illustrates the keto acid (KBA) component. Also shown is the measured vs. predicted concentrations from the LOOCV for the calibration data.	188
Figure 5-15 - (i) Regression coefficient 1 plotted alongside the pure component spectrum of ABA and KBA. (ii) Second regression coefficient plotted with pure component ABA and KBA spectra. (iii) Both regression coefficients associated with	

the constructed model superimposed. (iv) First regression coefficient and inverse of second regression coefficient superimposed.....	191
Figure 5-16 - Plot of absorbance measured against variable number for the variables identified by the genetic algorithm for inclusion in the PLS model.....	193
Figure 5-17 – Plots of measured concentration against predicted concentration for two validation datasets for the model built based on the first derivative regions identified by the genetic algorithm.	196
Figure 5-18 - Plots of measured concentration (mMol dm ⁻³) (from the reference analysis) against predicted concentration (mMol dm ⁻³) (from the PLS model) for the off-site replicate of the biotransformation process. (i) Refers to the amino acid (ABA) concentration whilst (ii) refers to the keto acid (KBA) concentrations.	200
Figure 5-19 – Scores associated with PC 2 plotted against PC1 scores for the mid infrared calibration dataset and the off-site process replicate.....	202
Figure 6-1 - De-racemisation of D/L-alanine using the DAAO enzyme contained within a <i>Pichia pastoris</i> host generating L-alanine and pyruvic acid.....	204
Figure 6-2 - Setup of bioreactor system illustrating the location of the mid infrared probe during the biotransformation process and the probe design with diamond ATR crystal.....	205
Figure 6-3 - Calibration curves for the quantification of pyruvic acid and acetic acid using the developed organic acids HPLC method.	208
Figure 6-4 - Pure component absorbance and first derivative mid infrared spectra of the three key analytes of interest in the biotransformation process acquired at-line using the mid infrared ATR probe system. Due to absorbance by the diamond ATR crystal only the region between 800 cm ⁻¹ and 1900 cm ⁻¹ was investigated.....	211
Figure 6-5 - Mid infrared spectra of biotransformation samples obtained at various resolutions to determine the optimum parameters for the spectral acquisition.	213
Figure 6-6 - TCA cycle adapted from Prescott <i>et al.</i> illustrating the routes by which pyruvic acid can be utilised either by conversion to acetyl-CoA or carbon fixation into the TCA cycle for energy generation.	217
Figure 6-7 - Oxidation of pyruvic acid to acetic acid, water and carbon dioxide by hydrogen peroxide. ¹⁴⁵	218

Figure 6-8 – Mean alanine concentration plotted against process time. Interpolation of the unknown values using a linear approach where a linear line connected the data points did not accurately represent the missing points. From the shape of the plot it was also clear that a linear trend was not the best fit for this dataset.	220
Figure 6-9 – Exponential decay fitted through the reference data points to give a more accurate interpolation of the alanine concentration at the time points for each spectral acquisition.	221
Figure 6-10 - Plot of natural log of alanine concentration against process time. Having determined the process proceeded via first order kinetics, the concentration at the time each spectrum was acquired was estimated based on the equation of the linear line of best fit.	222
Figure 6-11 – (i) Plot of scores associated with latent variable 1 plotted against scores for latent variable 2 and (ii) scores associated with latent variable 1 plotted against spectrum number to identify the five datasets to be used for calibration and the external validation dataset.	224
Figure 6-12 - Plots of measured concentration against predicted concentration. (i) Internal validation of PLS model constructed using the full spectral range acquired (800 cm^{-1} through 1900 cm^{-1}). (ii) External validation dataset alongside calibration data.	225
Figure 6-13 – Plots of measured alanine concentration against predicted concentration for PLS model constructed using refined spectroscopic regions. (i) Calibration data with predicted concentrations from internal validation procedure. (ii) External validation dataset shown alongside internal validation data.	228
Figure 6-14 – (i) Loadings associated with latent variable one of the constructed model plotted along with the pure component spectrum of alanine and (ii) loadings associated with latent variable two of the model plotted with the pure component spectrum of alanine.	229
Figure 6-15 - Regression coefficients associated with the constructed PLS model and infrared spectrum of pure component alanine for comparison.	230
Figure 6-16 - Plots of pyruvic acid and acetic acid concentration against process time for two biotransformation process replicates illustrating (i) the consistency in pyruvic	

acid concentration during the process and (ii) the consistent concentrations of acetic acid before accumulation towards the end of the process.	233
Figure 6-17 - Plots of first derivative value at 1245 cm ⁻¹ plotted against process time to illustrate the little overall variation observed in the spectral features over the course of the biotransformation process. This confirmed the reference analysis data suggestion that little or no variation in pyruvic acid concentration was observed during the process.	234
Figure 7-1 - Plot of scores associated with first principal component against process time for third replicate of fermentation process. Clearly identifies the initial growth phase, point of glucose limitation, feeding and finally the point of induction. Also shown is dissolved oxygen profile for fermentation, which confirms the point of glucose limitation by the increase in dissolved oxygen levels.	242
Figure 7-2 - Predicted classifications of the various stages in the fermentation process for the spectra from the fourth replicate of the process designated for validation purposes.	245
Figure 7-3 – Responses for the external validation dataset predicted using the PLS-DA model constructed for the fermentation stage of the process.	247
Figure 7-4 - Output from neural network for unseen fermentation replicate used to identify which phase in the fermentation each spectrum relates to.	249
Figure 7-5 - Output for the validation dataset from the neural network constructed using the mean centered second derivative near infrared spectra.	251
Figure 8-1 - Summary of the biotransformation process for the generation of compound B from S-alanine/L-alanine and compound A substrates proceeding via transaminase enzymes.	255
Figure 8-2 - Pathway illustrating the breakdown of polysaccharides to a monosaccharide and then to pyruvic acid and onto the TCA cycle adapted from Prescott <i>et al.</i> ¹⁵	256
Figure 8-3 – Selected features from NMR spectrum of compound B produced using the adapted procedure outlined by Clark <i>et al.</i> (2001) and purified by flash chromatography with a strong cation exchange resin.	260
Figure 8-4 - Calibration curves for the quantification of L-alanine and compound B using the developed HPLC method.	262

Figure 8-5 - Calibration curve for the quantification of compound A using the described HPLC method.	264
Figure 8-6 - Second derivative pure component spectrum of key analytes of interest in the first overtones region.	265
Figure 8-7 - Near infrared spectrum of water acquired using <i>in-situ</i> transreflectance probe at a variety of agitation rates likely to be experience during the biotransformation process to assess the effect on background noise in the spectra.	267
Figure 8-8 - First derivative spectrum of pure component samples of key analytes in the biotransformation process. Main region of interest shown between 950 cm ⁻¹ and 1760 cm ⁻¹	268
Figure 8-9 - Latent variable 1 scores plotted against latent variable 2 scores to identify the process replicates used in the calibration model thus ensuring the maximum variance observed was included in the calibration model.	271
Figure 8-10 - Plot of RMSEC and RMSECV values against latent variable number to determine the optimum number of latent variables for use in the constructed PLS model.	272
Figure 8-11 - Plots of predicted concentration (mMol dm ⁻³) (from PLS model) against measured concentration (mMol dm ⁻³) (from reference analysis) for the various key analytes of interest for both the internal (i, ii & iii) and external (iv, v, vi) validation procedures.	274
Figure 8-12 - Regression coefficients associated with the constructed PLS model for the biotransformation process.	276
Figure 8-13 - Plots of measured concentration (mMol dm ⁻³) (from HPLC methods) against predicted concentration (mMol dm ⁻³) (from PLS model) for the external validation dataset.	279
Figure 8-14 - PCA scores plot of all spectroscopic data obtained for the original samples and the retrospectively spiked samples.	280
Figure 8-15 - Comparison of first derivative mid infrared spectra of the samples identified using the leverage method acquired on the original and new ATR crystals.	285
Figure 8-16 - Scores plot from PCA carried out on the original data (original ATR crystal), the spiked datasets (new ATR crystal) and the transferred spectra. Transfer	

parameters were determined using the samples identified using the leverage method acquired on both the original and new ATR crystal.	287
Figure 8-17 - PCA scores plot test spectra transferred using multiple window sizes to determine the optimum parameters for the calibration transfer.....	291
Figure 8-18 - PCA scores of original samples after calibration transfer process compared against PCA scores for the spectra acquired using the original crystal and the spiked samples acquired using the new ATR crystal.....	291
Figure 8-19 - PCA of spectra acquired using original ATR crystal, new ATR crystal and SST transfer of the original crystal to the new crystal.....	293
Figure 8-20 - PCA carried out on original samples acquired on original crystal, spiked samples acquired on new crystal and the transferred samples converted from the original crystal to the new crystal.	294
Figure 8-21 - Plot of measured concentration (mMol dm^{-3}) against predicted concentration (mMol dm^{-3}) for the validation dataset with the PLS model constructed using spectra that had been transformed from the original crystal using the SST algorithm.....	296

List of Tables

Table 1 - Composition of salt solution, trace element solution and inoculum media used for the growth of <i>E-coli</i> strain.	85
Table 2 - Batch media composition used for the growth of <i>E. coli</i>	87
Table 3 – Gradient elution method used by industrial partner for the separation and quantification of the total amino acid content in the samples.....	100
Table 4 - Calibration data used to determine equation of linear line of best fit through the data as being $Y = 10429358.2x + 7892698$, and calculation of predicted values to determine RMSE.	104
Table 5 - Peak areas and calculated concentration of ten replicate injections to assess the repeatability of the developed method.....	105
Table 6 – Theoretical locations of the signals expected to be observed in the pure component spectra of the TBG and TMP analytes.	109
Table 7 - Approximate concentrations of TBG and TMP of the samples whose infrared spectra were superimposed to illustrate the variations observed at concentrations typically expected during the biotransformation process.....	113
Table 8 - Theoretical regions where signals from the TBG and TMP were expected to be observed.	116
Table 9 - Summary of key model parameters for the constructed PLS model for the biotransformation process and the external validation datasets.....	127
Table 10 – Summary of key model parameters and the external validation procedure for the model constructed using the refined spectral regions.	129
Table 11 – Error values associated with the PLS model constructed from the single spectra of the full fingerprint region and the mean concentration of the reference data.	130
Table 12 – Errors associated with the model constructed using the single spectra of the selected mid infrared regions and the mean values of the reference analysis. ..	132
Table 13 – Error values associated with the prediction of the two off-site biotransformation replicates.	145
Table 14 - Theoretical locations of signals expected to be observed for ABA and KBA.....	160

Table 15 - Initial concentrations of amino acid and keto acid determined by experimental design in an attempt to break any co-linearity within the system.....	166
Table 16 – Summary table of the errors associated with the calibration, internal and external validation for the constructed PLS models for the various identified spectral regions of interest.	168
Table 17 - Summary of error values for the calibration, internal and external validation procedures for the PLS model constructed from the raw mid infrared fingerprint spectra.....	182
Table 18 - Error values associated with the PLS model constructed from the first derivative fingerprint region mid infrared spectra.....	187
Table 19 - Error values for PLS model constructed using the genetic algorithm....	193
Table 20 - Errors of calibration and prediction associated with first derivative spectra based on regions determined by the genetic algorithm.....	195
Table 21 – Summary of the models constructed for the quantification of ABA and KBA during the biotransformation processes.....	198
Table 22 - Experimental design used to determine the initial concentrations of alanine and pyruvic acid added to each replicate of the biotransformation process to ensure the stoichiometric linkage was broken.....	215
Table 23 - Dry cell weights determined for samples removed during the fermentation process*.	239
Table 24 – Classification of each IR spectrum by assigning it to a stage in the fermentation process based on the scores associated with the first principal component.....	244
Table 25 - Number of principal components associated with the various stages of the fermentation process used to construct the SIMCA model for the fermentation. ...	245
Table 26 - Assignment of the selected signals in the acquired proton NMR of synthesised compound B for use as analytical reference material.....	261
Table 27 - Errors of calibration and cross validation associated with the calibration model along with the errors of prediction for the unseen validation dataset.....	273
Table 28 - Errors associated with the calibration model constructed using the spiked datasets and prediction of the key analyte concentrations in the external validation dataset.	278

Table 29 - Errors of calibration, cross validation and prediction associated with model constructed from both the spiked samples and the original samples obtained from the biotransformation process.	279
Table 30 - Samples identified using leverage method to be scanned on new ATR crystal to determine the calibration transfer parameters.	283
Table 31 - Samples identified using inverse method to be scanned on new ATR crystal to determine the calibration transfer parameters.	284
Table 32 - Errors associated with the prediction of concentration of the validation dataset using the direct standardisation transferred spectra.	288
Table 33 - Errors of calibration for the spiked PLS model and errors of prediction for the transferred spectra from the original crystal to the new crystal.	292
Table 34 – Errors of calibration, cross validation and prediction for the PLS model constructed using the SST transferred spectra and spiked samples. Validation utilised the SST transferred spectra for the fourth biotransformation replicate designated for validation.	295

Abbreviations

ABA	Amino butyric acid
ANN	Artificial neural network
ATR	Attenuated total reflectance
Conc ⁿ	Concentration
Cpd.	Compound
CSV	Comma separated variable
DAAO	D-amino acid oxidase
DAD	Diode array detector
DS	Direct standardisation
FAD	Flavin-adenin dinucleotide
FDA	Food & Drug administration
FT	Fourier transform
FT-IR	Fourier transform infrared
SPC	Galatic file format
GC-MS	Gas chromatography mass spectrometry
HPLC	High performance liquid chromatography
KBA	Keto-butyric acid
LOOCV	Leave one out cross validation
LC-MS	Liquid chromatography mass spectrometry
MIR	Mid Infrared
MLR	Multiple linear regression
MSC	Multiplicative scatter correction
NIR	Near Infrared
NMR	Nuclear magnetic resonance
OPA	<i>Ortho</i> -phthalaldehyde
PLS	Partial least squares
PLS-DA	Partial least squares - discriminant analysis
PDS	Piecewise direct standardisation
KBr	Potassium bromide
PCA	Principal component analysis

PCR	Principal component regression
PAT	Process analytical technology
QbD	Quality by design
RMSE	Root mean square error
RMSEC	Root mean square error of calibration
RMSECV	Root mean square error of cross validation
RMSEP	Root mean square error of prediction
SNR	Signal to noise ratio
SIMCA	Singular isolinear method of class assignment
SST	Spectral space transformation
SNV	Standard normal variate
Std Dev	Standard deviation
STR	Stirred tank reactor
TBG	Tertiary butyl glycine
TFA	Tri-fluoro acetic acid
TMP	Tri-methyl pyruvic acid
ZnSe	Zinc selenide

Abstract

Industrial biotransformation processes are becoming increasingly important for the production of single enantiomers of both low value commodity and high value fine chemicals. Despite this demand and the regulatory authorities encouragement of a quality by design approach, the application of process analytical technology to these systems has, to date, been relatively limited. A more traditional off-line approach involving chromatographic methods is still commonly employed for the quantification of key analytes during the process. *In-situ* measurements tend to be limited to physical parameters of the system such as pH and dO_2 , which give little information about the actual process progression.

This study investigates the potential of applying infrared spectroscopic techniques to monitor and quantify the key components of de-racemisation and transaminase biotransformation processes. Multivariate models based on the near and mid infrared spectroscopic regions have been constructed for a variety of these processes. Each constructed model was subjected to an external validation procedure to ensure rigorous testing. Stoichiometric linkages were known to exist within these systems. Whilst steps were taken to ensure these linkages were broken, the contributors to each model were also carefully examined to ensure that co-linearity within the constructed models had been adequately addressed. Having constructed robust process models, mechanisms of ensuring the long-term suitability of the models were also investigated. This aimed to ensure the continued predictive ability of the constructed models following instrument maintenance, repair or replacement.

Quantitative models resulted that were able to predict the key analyte concentrations of the external validation datasets over the course of the biotransformation processes. Predicted values from the constructed models were in good agreement with both the errors of calibration and cross validation associated with the models, and the actual concentrations predicted by the off-line chromatographic reference methods.

1. Literature Review

1.1 Introduction

Currently there is a demand for the manufacture of chiral compounds for use in the manufacture of pharmaceuticals. The stereochemistry of molecules has proven implications on drug efficacy, as well as other features such as drug metabolism and excretion of the compound from the body. Chirality has proven such an important factor in drug development that many pharmaceuticals currently produced in racemic form are being re-examined to determine if single isomeric forms provide more effective treatments.¹

Amino acids exhibit chirality and are rapidly becoming important intermediates in the synthesis of many pharmaceutical compounds as well as other industrially relevant compounds. However mechanisms for the medium to large scale (10-15 L) production of such un-natural chiral amino acids are limited.²

Fotheringham *et al.* (2006) reported the development of a general bio-catalytic approach to the production of multiple unnatural chiral amino acids in high enantiomeric excess. The approach selectively converted the undesired enantiomer of a racemic amino acid mixture using D-amino acid oxidase enzymes expressed by a microorganism.² These DAAO enzymes are naturally occurring within mammalian cells (pig kidney) and in yeasts.^{3, 4} The enzymes can however be expressed by recombinant *Escherichia coli* cells and used for the production of unnatural amino acids and α -keto acids.⁵

As an example of the generation of such unnatural chiral amino acids, Fotheringham *et al.* (2006) cited the preparation of L-2-aminobutyric acid in enantiomeric excess of over 99% from the inexpensive, racemic 2-aminobutyric acid substrate.²

Within the pharmaceutical industry regulatory authorities are actively encouraging the implementation of technology that promotes greater understanding and control of critical process parameters in the manufacture of drugs and drug intermediates. Support of this approach culminated in the 2004 publication of the process analytical technology (PAT) directive by the United States Food and Drug Administration (FDA).⁶

Such a lucrative process for the generation of these important chiral amino acids provided the ideal opportunity to investigate the implementation of PAT, particularly the already well established technique of infrared spectroscopy, for monitoring and control of these industrially relevant biotransformation processes. Adopting such a methodical and structured approach will allow for greater process understanding and improved process control. As a consequence this will result in an improved product quality, potentially leading to improved yields and better final product quality.⁷ It was however also synonymous with the regulators insistence on the generation of high quality products consistently.⁶

1.1.1 Process Analytical Technology Initiative

The key driving force behind the implementation of PAT to bioprocesses, particularly those employed for the production of high value recombinant protein based drugs, was the publication in 2004 of the United States Food and Drug Administration PAT initiative.⁶

This initiative signified a shift in position from the traditional approach of testing the final product of a process for quality and rejecting batches that failed to meet these requirements.⁸ Going forward a “quality by design” (QbD) approach to manufacturing was suggested for implementation. This approach centred around introducing the concept of process understanding, control and consistency to not only improve the quality of the end product but ensure consistency of quality.⁹ A philosophy concisely expressed within the published recommendations:

“...quality cannot be tested into products; it should be built-in or should be by design...”

U.S Department of Health & Human Services, Food & Drug Administration (2004)⁶

In order to attain such quality and consistency it was necessary to have a thorough understanding of the process, deviations within the process and also any interactions that can and do occur during the process. Adopting a QbD approach shifted the focus firmly towards implementing sound scientific methodology throughout the process from raw materials, through manufacture and onto products.⁸

The detailed understanding of such processes is attained using PAT analytical techniques, which incorporate chemical, physical, biological and mathematical approaches to monitoring the system.⁶ These measurements can then be used for feedback control to maintain the process consistently within acceptable parameters thus ensuring the quality of the product.⁹

As well as ensuring the quality of the final product there are a number of advantages to the manufacturer of adopting a QbD approach and the recommendations set down by the PAT initiative. A reduction in the amount of rejected product would be an obvious advantage of the system. Any reduction in wastage or improvement to final product quality would result in financial advantages to the manufacturer. Streamlined and better-understood processes were also a likely outcome of implementation of the PAT initiative ultimately resulting in decreased costs and increased product yields.^{6,9}

1.1.2 Bioprocesses

The term bioprocess is broad and can be used to refer to multiple biological reactions or processes. It can be used to describe the cultivation of cellular material (a submerged culture process), the production of recombinant products (natural or foreign proteins) or the catalytic conversion or generation of a product (biotransformation).¹⁰ Processes involving the use of biological catalysts have

become increasingly common in the industrial setting but also at the laboratory development scale.¹¹

1.1.2.1 Chemical Synthesis vs. Biocatalysis

Details of the mechanisms and routes utilised in the generation of compounds adopting a synthetic chemical approach are well understood, with a substantial knowledge and experience base available.¹² Whilst bio-catalytic routes may not be as well understood they offer a number of key advantages over chemical approaches. High levels of regio-selectivity and stereo-selectivity can often be achieved in fewer stages and under milder reaction conditions than the corresponding chemical approach. Clearly bio-catalytic processes offer a number of distinct advantages over traditional chemical approaches, however there are a number of important factors that must be considered to determine if they are indeed the best approach.^{12, 13}

1.1.2.1.1 Process Considerations

A number of key issues, specifically biocatalyst availability, purity and the associated costs of production, both in terms of substrate and downstream processing, need to be considered before utilising these biocatalysts. The economics of the process have meant that bio-catalytic processes have been limited to the production of high value specialty chemicals.^{11, 12} However some common industrial processes do employ biocatalysts, specifically the generation of acrylamide from acrylonitrile and the generation of fructose from glucose.¹¹

Whilst there are some challenges associated with the utilisation of biological catalysts these can be outweighed by the favourable process conditions. Generally biologically catalysed processes can be carried out under much milder conditions (temperature, pressure and pH) compared with synthetic organic processes. Since the main solvent utilised is water, these bioprocesses also have 'green' credentials compared with the more traditional approaches.¹³

1.1.2.1.2 Bioreactor

The biocatalysts that are to be exploited are naturally occurring enzymes commonly found in plant, animal or microbial cells. To obtain these proteins in sufficient quantities microbial cells, such as *Escherichia coli* (*E. coli*) can be modified. Modification involves the insertion of a plasmid into the host cell, that upon induction produces large quantities of the desired enzyme.⁴

Growth of these organisms to produce sufficient levels of biomass, and therefore protein, would require the use of a suitable bioreactor. Whilst the stirred tank reactor (STR) design is regarded as the standard¹⁴ there are some instances where other more specialised bioreactor designs may be more suited for the catalytic process. An example of such a process is the production of tertiary-L-leucine using leucine dehydrogenase enzymes. The process requires the regeneration of a co-factor, a molecule that binds to a protein bringing about biological activity, and is better achieved using a membrane reactor rather than the standard STR design.¹²

1.1.2.1.3 Enzyme Form

The form in which the biocatalyst is present in the system is also a key factor influencing the process. Whole cells can be utilised or alternatively the desired enzyme may be isolated as the free enzyme or immobilised onto an insoluble support structure. Various approaches to immobilisation have been reported, however the ultimate aim is to fix the catalyst in a particular location to improve turnover.¹³ Isolation or immobilisation of the enzymes would result in an increase in the costs associated with the bioprocess, as well as increasing the number of steps in the process.¹²

1.1.2.2 Industrial Applications

Despite their increased complexity, both in terms of the genetic modification required for expression of the enzyme and the conditions required to generate

sufficient quantities of the enzyme of interest, bio-catalytic processes have been utilised on the industrial scale.¹¹

Under the more specialty chemicals banner bio-catalytic processes have been utilised for the generation of chiral alcohols and D-amino acids. Biological pathways have also been reported for the generation of natural products from relatively cheap substrates such as the conversion of D-glucose to vanillin.¹¹

The application of bioprocesses on a larger industrial scale (production of ~250,000 L per day) has been more limited presumably due to the high production costs and low value product. Some processes are economically feasible such as the conversion of acrylonitrile to acrylamide and the breakdown of lactose to glucose and galactose for use in low-lactose milk.¹¹

1.1.3 Biotransformation Process

Many of the bioprocesses discussed fall into the category of biotransformation processes. By definition a biotransformation process brings about a minor, but highly specific, change to the structure of a molecule. These changes are carried out by a microorganism that is not undergoing growth. One of the key advantages is the control over stereochemistry offered by these processes in comparison with traditional chemical approaches.¹⁵

1.2 Spectroscopic Approaches to Process Monitoring

Vibrational spectroscopic (infrared) techniques offer a fast and non-destructive method of acquiring sample information.¹⁶ Unlike many traditional approaches to monitoring key parameters of a system, spectroscopic methods are able to monitor and, with the aid of chemometric models, quantify these multiple components in a system.^{17, 18}

The particular bioprocess of interest in this study was a biotransformation for the de-racemisation of an amino acid. Application of spectroscopic approaches to monitor such processes has been relatively limited; however there has been substantial application of the technique to other bioprocesses. Many of the challenges encountered in applying this technique in such systems, and the solutions, were likely to be consistent with those encountered with the biotransformation processes. For this reason the literature survey initially considered all bioprocesses before specifically focusing on biotransformation systems.

A variety of different approaches to the acquisition of infrared spectroscopic data have been reported. With spectroscopic measurements the objective was aimed less towards obtaining information about physical parameters of the system (such as pH and growth rate) and more related to deriving information regarding changes to key substrates or products.¹⁹

Sampling approaches can be classified into three major categories: off-line, at-line and on-line (Figure 1-1).

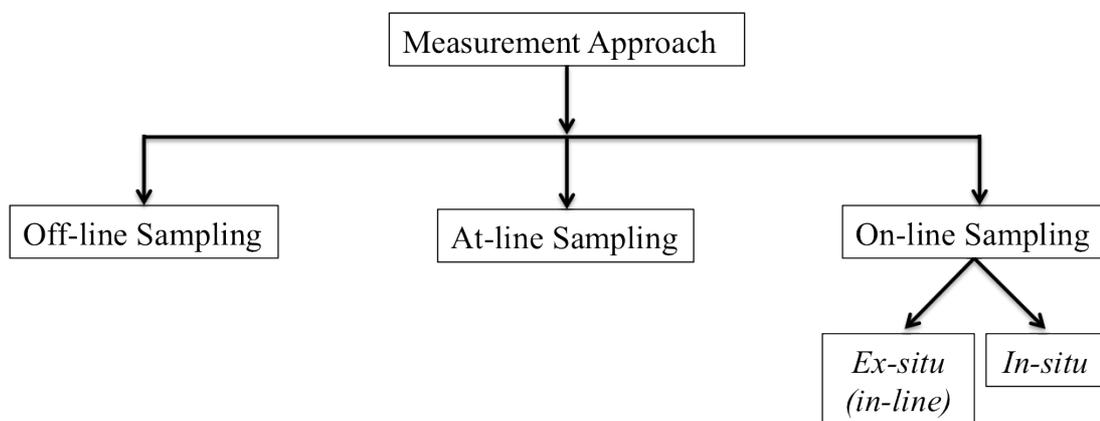


Figure 1-1 – Graphical representation of the various sampling approaches (Vaidyanathan *et al.* (1999)) utilised for the acquisition of the infrared spectra of the reactor contents at various stages during a bioprocess.¹⁹

As the name suggests, off-line sampling is an invasive process that involves the removal of a sample from the reactor. These samples are then stored, or in some

cases undergo sample pre-treatment, and analysed away from the location of the bioreactor.¹⁹ The key issue with this approach is the substantial time lag that is introduced from the point of sampling until the sample information is reported. From a process control perspective, particularly of batch processes, this is not an ideal situation as in many cases the window of opportunity for corrective action will have passed by the time the relevant information has been obtained.

The at-line sampling approach has a number of similarities with the off-line method. Representative samples are removed from the process and analysed using instrumentation in the vicinity of the reactor.^{16,20} A rapid at-line sampling approach offers many advantages over the off-line method, however the process is still invasive with a time lag between sampling and the process information being available.

On-line sampling approaches actually take the sampling, in this case spectral acquisition, into the process, eliminating the invasive sampling procedure required by off-line and at-line methods.¹⁷ Located within, or as part of, the reactor vessel it also means a higher sampling frequency allowing for the possibility of obtaining near real time process information.¹⁹ The class can be further sub-divided into *ex-situ* and *in-situ* sampling mechanisms.

Ex-situ (in-line) approaches place the analyser outside the actual reactor vessel with an aliquot of the process fluid diverted through a flow cell structure to allow the required measurements to be made. Once the various measurements have been made the extracted fluid is then returned to the reactor.¹⁹⁻²¹

Alternatively an *in-situ* approach to sampling may be taken. This places a probe inside the reactor system and in contact with the process fluid. Measurements made in this manner are regarded as the optimum sample collection mechanism. This non-invasive sampling approach is often regarded as the ideal solution to near real time process monitoring.¹⁹⁻²¹

Even although a variety of sampling approaches for the acquisition of spectroscopic data were available, a clear list of preference had emerged. However the feasibility of implementation would depend on the spectroscopic region being investigated and the availability of the necessary equipment.

1.2.1 Near Infrared Spectroscopy

Near infrared spectroscopy (NIR) has been employed in a variety of industrial settings ranging from the food and agricultural industries through to the more recent move to into the pharmaceutical and petrochemical industries.²² The near infrared spectroscopic region covers the wavelength range from 700 nm through to 2500 nm (4000 cm^{-1} to 14285 cm^{-1}).²³ Signals in this region of the electromagnetic spectrum are a result of overtones and combination bands of the fundamental bending and stretching of bonds observed in the middle (mid) infrared region.^{23, 24} These transitions are however formally forbidden but can result from the molecules behaviour as an anharmonic oscillator (discussed further in section 2.4). As a result of this the vibrational transitions do not occur frequently so the resulting signals are generally very weak, reportedly between 10 and 100 times weaker, some suggest even 1000 times weaker²⁰, than a corresponding allowed transition in the middle infrared region.²⁴ Signals in this region also tend to be broad and overlapping and require chemometric techniques to elucidate spectral features.²⁵

These weak, broad and overlapping signals may appear problematic from a theoretical perspective, however when attempting to monitor biological systems in particular these features are actually advantageous.

Weaker signals will allow for the use of longer path lengths compared with the traditional small optical path lengths associated with other spectroscopic approaches. Longer path lengths subsequently allowed for the analysis of a larger sample volume thereby giving greater confidence in the sample homogeneity.¹⁸ By extension therefore this would imply that analytes present at relatively low concentrations could also be monitored.

By their nature, biological processes are optically challenging matrices. Highly turbid, light scattering bacterial systems or clean but complex cell culture media containing many correlated analytes, both benefit from this weak nature of the signals in the near infrared region.²⁴⁻²⁷

1.2.1.1 Spectral Acquisition

Light incident on the presented sample can be measured in three ways; the transmitted light, reflected light or via a hybrid method based around transreflectance (Figure 1-2). As the name suggests transmission spectroscopy directs the incident beam of light through the sample and then measures the light that has passed through the sample. Diffuse reflectance measurements direct an incident beam of light onto the sample and measures the light reflected back by the sample. Transreflectance measurements are a hybrid of the two approaches and are discussed in more detail when considering near infrared probe designs for *in-situ* applications (1.2.1.5).

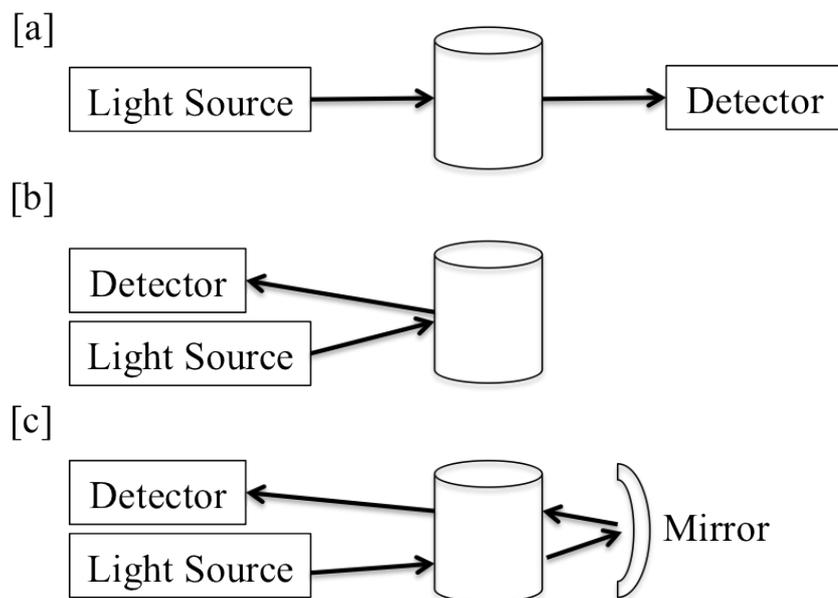


Figure 1-2 – Illustration of the three possible infrared measurement approaches: (a) transmission, (b) reflectance and (c) transreflectance. Adapted from Vaidyanathan *et al.* (1998).²⁰

The acquired spectra are commonly represented as absorbance spectra, however it is actually the unabsorbed light (transmitted or reflected) that is measured by the detector system. These measurements are then converted using Equation 1 (where T is transmitted light) or Equation 2 (where R is reflected light) depending on the collection mechanism, to give the familiar absorbance spectrum commonly associated with infrared spectroscopy.²¹

$$Absorbance = \log\left(\frac{1}{T}\right) \quad \text{Equation 1}$$

$$Absorbance = \log\left(\frac{1}{R}\right) \quad \text{Equation 2}$$

Although near infrared spectroscopy has become a popular, fast and non-destructive process analysis technique²³, there are some fundamental issues with the technique that must be taken into account. With biological systems in particular the background matrix will undoubtedly be mainly water.

The presence of water in the background matrix can be problematic due to the strong absorbance exhibited by the OH vibrations. Techniques such as employing a shorter sample path length, increasing the number of averaged scans or using mathematical algorithms often need to be employed to reduce the dominance of these effects.^{22, 28}

1.2.1.2 Near Infrared Instrumentation

Within the near infrared region two spectrometer designs dominate – the dispersive instrument and a Fourier transform instrument. Both types of instrument are commonplace, although they are optically very different. Further details on the optical differences between the two systems are discussed in section 2.4.3.

Briefly dispersive instruments utilise a series of prisms, mirrors and grating monochromators to separate the light source into the various component wavelengths. By moving the position of the mirrors and gratings the infrared spectrum at the various wavelengths can be obtained. Fourier transform infrared (FT-IR) instruments

utilise a Michelson interferometer to separate the light into its component frequencies and generate an interferogram. A Fourier transform algorithm is then used to convert the interferogram to the traditional infrared spectrum.²¹ FT instruments offer a number of advantages over the more traditional dispersive design, with an improved signal to noise ratio being one of the key advantages.^{21, 29}

1.2.1.3 Off-line/At-line Sampling

In the vast majority of cases at-line near infrared samples were presented to the spectrometer system for analysis in quartz cuvettes. This approach provides a quick, easy and convenient method of altering the sample path length, typically between 0.5 mm and 2 mm. In most reported cases involving at-line sampling with cuvettes the acquired spectra have been the transmission spectra.^{16, 22, 24-26, 30-32} Presentation of the sample in cuvettes is not restricted to transmission measurements. Reflectance spectra of samples can also be obtained using a cuvette based sample presentation approach. Typically however these cuvette path lengths tend to be substantially larger, with path lengths of 1 mm up to 10 mm having been reported.^{33, 34}

An adaptation on the cuvette based sampling approach introduces temperature control to the system. The sample is still presented to the spectrometer in a quartz cuvette however the sample chamber is temperature controlled. Adopting this approach will reduce any temperature dependent spectral variations improving the reproducibility of the spectra.²⁸ In addition this approach allows the sample spectra to be acquired under conditions that were consistent with those in the bioreactor.^{27, 35-38}

1.2.1.4 In-line Sampling

The in-line sampling approach represents a method of on-line monitoring but in an *ex-situ* manner as the spectral acquisition device is located outwith the reactor system. Compared with at-line and *in-situ* sampling methods there has been relatively limited application of in-line sampling and spectral acquisition approaches with near infrared

systems. Particularly with biological systems in-line sampling approaches have been prone to spectral artifacts, dead volumes and sterility problems.¹⁷

Generally in-line sampling systems utilise a flow through cell design that mimics the quartz cuvettes used with the at-line sampling approach. With the aid of a pump the sample material is removed from the bioreactor system, passed through the flow cell where the near infrared spectrum is obtained before being returned to the bioreactor system (Figure 1-3). Sample intervals for this method have been reported at thirty-minute intervals up to twelve hours.^{39, 40}

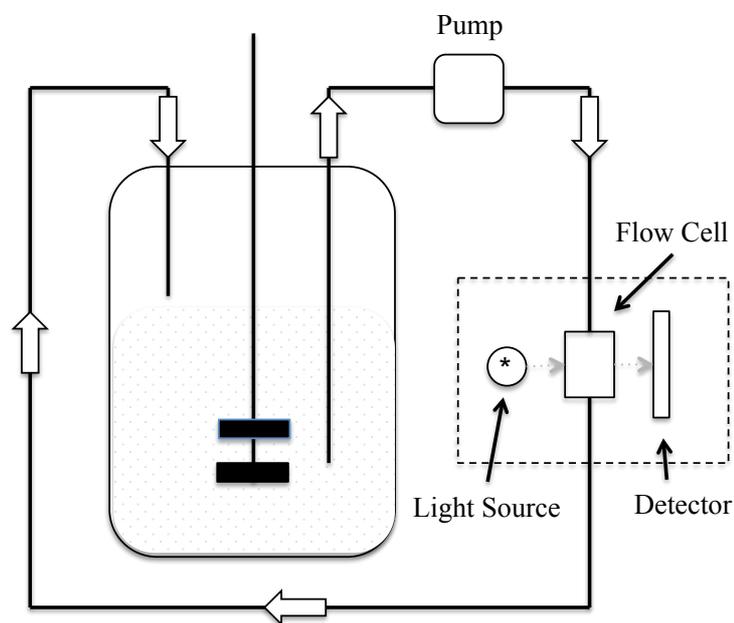


Figure 1-3 - Schematic representation adapted from Holm-Nielsen *et al.* (2007) of typical in-line flow through system employed for sampling and spectral acquisition in the near infrared region.³⁹

An alternative approach to in-line sampling was put forward by Ge *et al.* (1994). This approach utilised a custom-built infrared spectrometer system with the light transmitted via a fibre-optic bundle. The tip of this bundle was placed in contact with the glass walls of the bioreactor system and spectroscopic measurements made based on the diffuse reflectance principal.⁴¹ This setup was essentially similar to using a reflectance probe located outwith the reactor to make the required measurements. As

the technology now exists to allow this approach to be readily applied *in-situ* there may be limited applications for this particular setup, unless dictated by an external factor such as reaction conditions.

Although possible in-line sampling approaches for the near infrared region in particular have been relatively limited, mostly likely due to the aforementioned difficulties with the technique (sterility and spectral artifacts)¹⁷ and that robust *in-situ* sensors are readily available.

1.2.1.5 In-situ Sampling

Sampling approaches applied *in-situ* are achieved using probe based sampling systems that place the 'sample window' in physical contact with the sample matrix inside the bioreactor system.^{21, 23, 25}

Within the NIR frequency region, with the exception of the combination bands, light can be transmitted both to and from the spectrometer system using silica based fibre-optic cabling.¹⁷ This fibre optic transfer cable can consist of either a single fibre to transfer the light to the sample and a single fibre to carry the light back to the detector or a bundle of fibres in each case. The type of measurement being made may however dictate which approach was the more suitable, with reflectance measurements favouring a fibre bundle due to the greater probabilities of scattered light being collected.⁴²

A variety of different probe designs based on the three key measurement principals (Figure 1-2) are available, highlighting the flexibility offered by the technique.²⁰

In-situ transmission measurements of a process sample can be made relatively simply. An optical fibre conveys the light from the radiation source to the sample. The light then passes through the sample at a fixed path length before being collected by a second fibre at the opposite end of the sample material and transported to the detector system.⁴³ By placing two fibres opposite each other, at a fixed distance apart

within the process environment, and encasing in a metal housing an *in-situ* transmission probe can be constructed. The sample can then be passed through the fixed gap allowing acquisition of the spectra.⁴⁴

In situations where the sample matrix contains a high concentration of particulate materials or the matrix is highly light scattering, diffuse reflectance measurements can provide the best quality spectrum.⁴⁴ Light from the source is directed via optical fibres to the sample location. Any light reflected by the sample is then collected by another fibre bundle and transported to the detector system.⁴⁵

Probe systems based on the transmittance measurement principal have been commonly utilised with biological systems.^{23, 46-49} In this case the sample is placed onto a reflective surface so the incident beam of light passes through the sample material (as in transmission spectroscopy) is then reflected off the mirror and back through the sample material a second time before passing to the detector (Figure 1-4).⁵⁰ Variations to the path length through which the light travels can then be made either by altering the amount of sample material present (size of sample gap) or the angle at which the incident light is directed to the sample. Depending on the nature of the presented sample the acquired spectra may also contain an element of diffuse reflectance that must also be taken into account.^{42, 43}

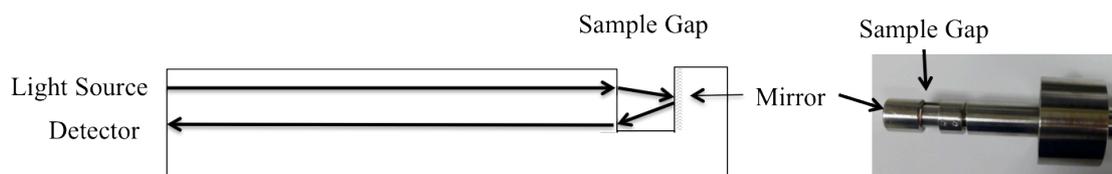


Figure 1-4 - Schematic representation and image of transmittance probe for use with near infrared spectroscopy adapted from Roychoudhury *et al.* (2007), Vaidyanathan *et al.* (1998) and Hassell *et al.* (1998).^{20, 42, 46} Key components of the probe have also been indicated.

1.2.2 Middle Infrared Spectroscopy

The middle infrared (MIR) spectroscopic region incorporates the frequency range from 400 cm^{-1} through to 4000 cm^{-1} .^{19, 51} Unlike the near infrared, signals in the mid infrared region arise due to the fundamental vibrations of the bonds within a molecule. Under the rules defined by quantum mechanics these are allowed vibrational transitions and so can occur easily and frequently. The signals observed are therefore generally much more intense than those obtained within the near infrared region.^{52, 53}

Signals in the mid infrared region also tend to be sharper and more resolved than are observed within the near infrared region. Within the fingerprint region (below 1200 cm^{-1}) unique, analyte specific patterns are recorded in addition to the useful features observed over the remainder of the region. These unique patterns can therefore be utilised for structural elucidation and identification of unknown compounds as well as the quantification of known analytes of interest.^{51, 53-56}

Despite one of the key positives of mid infrared being that the signals arise due to fundamental, allowed transitions this is also one of the greatest challenges associated with the technique. Since the transitions are allowed they occur readily and essentially cause saturation of the detector. In order to acquire an acceptable mid infrared spectrum very small sample path lengths, or small sample amounts, are required. Traditional off-line laboratory methods such as the preparation of potassium bromide (KBr) discs or Nujoll mulls achieve this, however these approaches cannot be easily transferred to in-line or even *in-situ* approaches.⁵⁴

In a related issue the strong absorption characteristics of water, a key component in many biological fluids, could dominate the spectra in a phenomenon similar to that observed with the near infrared region.^{55, 57}

In order to overcome these issues it was necessary to utilise a sampling approach that allowed for these essential short path lengths (typically between 10 and $100\text{ }\mu\text{m}$ ⁵⁸) to

be obtained but still allow for rapid sampling approaches essential for process analysis.

Suitably small path lengths (in the approximate region of 25 μm) could be obtained using attenuated total reflectance (ATR) crystals.^{54, 55, 58} This sampling approach places the ATR crystal surface in direct contact with the sample material, with the spectrum of the sample acquired as a result of an evanescent wave.⁵⁸

When light travelling through a medium with a particular refractive index (n_1) meets a boundary with a second medium of different refractive index (n_2) the beam of light can either undergo refraction or reflection. If the angle at which the incident beam of light hits the boundary is below the critical angle (determined by Equation 3) then the beam of light is refracted, where the angle through which it is refracted can be determined using Snell's law (Equation 4) and illustrated in Figure 1-5.⁵⁹

$$\theta_c = \sin^{-1} \left(\frac{n_1}{n_2} \right) \quad \text{Equation 3}$$

$$n_1 \sin \theta_1 = n_2 \sin \theta_2 \quad \text{Equation 4}$$

In cases where the light's angle of incidence is greater than the critical angle the beam of light will undergo total internal reflectance, where the angle of incidence is equal to the angle of reflection according to the law of reflection (Figure 1-5). On a related note it is this property of total internal reflection that allows light to travel through the fibre optic bundles common on many near infrared spectrometer systems.⁵⁹

A phenomenon that results from these total internal reflections allows the acquisition of an infrared spectrum using ATR crystals. When the incident beam of light is internally reflected at the crystal/sample interface an evanescent wave propagates into the sample medium. IR active groups within the sample can absorb energy from this evanescent wave resulting in the generation of an infrared absorbance spectrum based on the transmitted frequencies as before (Equation 1).⁵⁸

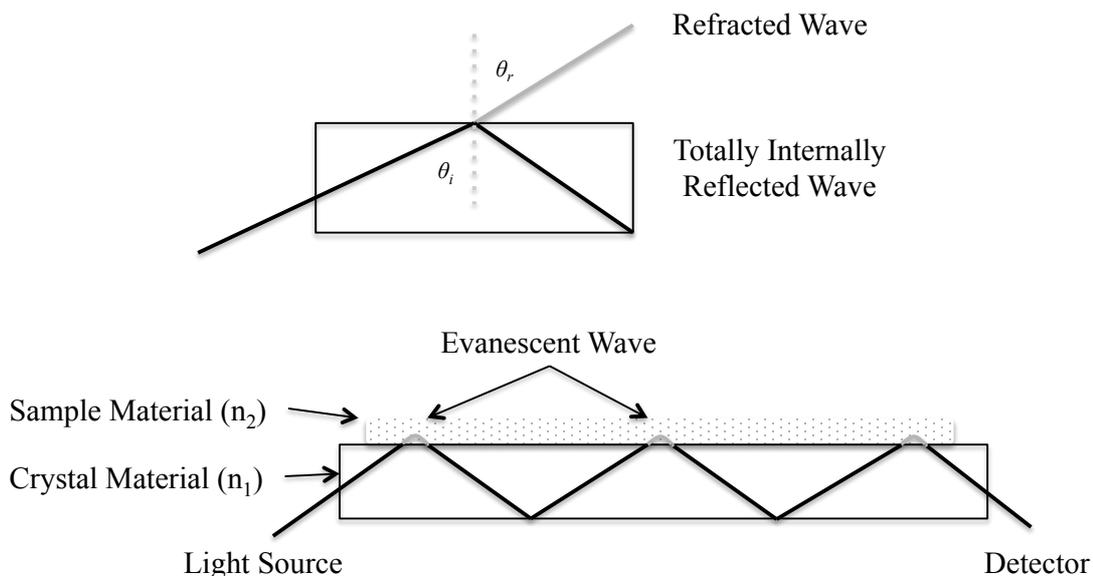


Figure 1-5 – Illustration of refraction and total internal reflectance of light at the boundary of materials and propagation of light through an ATR crystal illustrating the evanescent wave passing into the sample. In this case n_1 refers to the refractive index of the crystal material, n_2 the refractive index of the sample material, θ_i the angle of incidence and θ_r the angle of refraction.^{58, 59}

Since the evanescent wave only penetrates the sample material by a small amount this equates, over the multiple internal reflections, to an effective path length in the region of up to 25 μm . Ideal for the small path lengths required to acquire a suitable mid infrared spectrum of the sample material.⁵⁸

1.2.2.1 Off-line/At-line Sampling

The more traditional approaches of producing potassium bromide discs of the sample material or the preparation of a Nujoll mull could be applied either off-line or at-line.⁵⁴ These approaches however are time consuming, involve extensive sample preparation procedures and may not be well suited for the direct analysis of process samples, particularly biological matrices.

ATR crystal technology has been widely utilised for the at-line acquisition of mid infrared spectra.^{52, 53, 60} These systems house the ATR crystal within the spectrometer

design or via a sampling accessory. Different ATR crystal materials have been reported, the most common being zinc selenide (ZnSe) and diamond.^{53, 61}

Undoubtedly the associated cost implications will limit the available crystal surface area over which the sample can be placed. Diamond ATR crystals present a very small sampling surface however multiple internal reflections still occur resulting in an effective path length in the region of 10 μm .^{52, 53} Zinc selenide crystals offer a greater sampling surface area compared with diamond. This has the advantage that the crystal can be sunk into a trough allowing for easier liquid sampling and larger sample volumes.⁶²

ATR technology has also made the direct at-line sampling of solid materials possible, however good contact between the solid sample material and crystal surface is required to attain suitable spectra.⁵⁸

Although different crystal materials are available the most suitable will depend on the sample being analysed. Zinc selenide crystals have been reportedly damaged by samples containing acids or oxidising agents. Diamond ATR crystals appear to be the most chemically robust however the main disadvantage is their reduced optical range compared with ZnSe.⁵⁴

1.2.2.2 In-line Sampling

The in-line sampling approach has been more extensively adopted to acquire mid infrared spectra compared with uptake in the near infrared region. Setup of the system adopted the same basic principal as that described for in-line near infrared sampling³⁹, however due to the small path lengths required the sample acquisition mechanism was different.

Franco *et al.* (2005) reported an application of the traditional in-line approach where the sample fluid was directed through a short path length (25 μm) barium fluoride

flow cell.⁶³ This approach appeared to be in the minority, with most common examples employing ATR technology in place of the traditional flow cell.

In most situations the in-line flow cell was simply replaced by an ATR crystal over which the sample media was directed to acquire the infrared spectrum.^{55, 57} A slight adaptation of the approach was put forward by Acha *et al.* (2000) which aimed to reduce the dominating spectral features arising from the strong absorptions of water in the background matrix.⁵⁷ This was achieved by coating the surface of the ATR crystal with a permeable hydrophobic membrane. Analyte molecules were able to diffuse through the membrane layer to the crystal surface where they could interact with the evanescent wave. Water molecules were unable to permeate this membrane and so did not have an influence on the infrared spectrum acquired.

With small volume bioreactor systems an in-line flow through system was reported where an *in-situ* mid infrared probe was mounted in the flow cell.⁶⁴ Such an approach may have arisen due to the size of the bioreactor being unable to accommodate the mid infrared probe. However if the technology was available to take the measurements *in-situ* many of the issues associated with the in-line sampling approach could be avoided.

1.2.2.3 In-situ Sampling

With ATR crystals able to provide the very short path lengths required to obtain suitable mid infrared spectra the limiting factor in application of the technique *in-situ* has been connecting the probe to the spectrometer. Silica based optical fibres, such as those used with the near infrared systems, do not transmit light with wavenumbers below 3331 cm^{-1} (above 3000 nm or 3 μm).⁶⁵

Doak *et al.* (1999) reported the use of an *in-situ* ATR probe system to monitor an *E. coli* fermentation process.⁵⁶ This particular system did not employ optical fibres to transmit light from the sampling tip to the spectrometer but a series of 'knuckles' each of which contained a mirror to direct the radiation beam. Whilst being an

effective approach, its application in an industrial setting would be expected to be fairly limited due to a lack of flexibility and the close proximity between the reactor and spectrometer that would be required.

Developments in optical fibre design meant that light within the mid infrared region could be transmitted using fibres constructed from chalcogenides and metal silver halides.^{65, 66} These developments then allowed for the ATR crystal based sampling probe to be connected to the spectrometer system in a similar manner to the already well established near infrared systems.

As with the at-line and in-line approaches, a variety of different ATR crystal materials have been utilised with *in-situ* probe designs. Sapphire ATR crystal probes prevent access to the information rich fingerprint region (below 1200 cm⁻¹) due to absorbance of the crystal in this region.⁶⁷ Likewise diamond ATR crystals absorb light in the region around 2000 cm⁻¹ preventing use of this region.⁵⁴ Probes equipped with ZnSe ATR crystals have also been utilised, however these crystals have been reported to be incompatible with strong acids and some biological samples.⁵⁴

As with the other sampling approaches that employed ATR crystals, most applications utilised a diamond crystal as this appeared to be the more robust and chemically stable of those available, whilst still offering a suitably diverse spectral range.^{56, 67-69}

1.3 Process Analytical Technology

The application of PAT to conventional and novel bioprocesses has been reported.^{17, 18, 28, 44, 53, 70-73} As previously mentioned, although the main focus of this research was to investigate the feasibility of applying the technique to biotransformation processes, many of the discussions and challenges reported for other bioprocesses will be applicable to these biotransformation systems.

1.3.1 Submerged Culture Bioprocesses

1.3.1.1 Conventional Approaches

Typically the progress and understanding of bioprocesses have been determined by making measurements of critical process parameters such as biomass, nutrient or product concentrations.³⁰ Other features of the process such as pH, dissolved oxygen levels, temperature and substrate or product concentrations are all critical process parameters that require monitoring and control to achieve successful and reproducible processes.⁷⁴

Traditionally measurements of physical characteristics of the system, such as pH, dissolved oxygen and temperature have been made using *in-situ* probe systems that provide near real-time measurements of these critical parameters.^{30, 36} Whilst useful from a process control perspective these measurements give little information on how the bioprocess is actually progressing.

Indicators of process progression, typically biomass and nutrient levels, have been restricted to off-line measurement approaches.³⁰ Biomass measurements typically involve determining the dry cell weight of a medium aliquot or estimation based on the optical density measured at a wavelength of 600 nm.²³⁻²⁵ Key substrate and product concentrations can be determined using a number of approaches. Enzymatic assay kits have been reported for the quantification of a variety of typical analytes of interest (glucose, lactose, ammonia, glutamine and glycerol).^{26, 30, 49} Various chromatographic techniques such as gas chromatography (GC) and high performance liquid chromatography (HPLC) have also been employed as reference analysis methods.^{39, 47, 48} Each of these approaches requires substantial sample preparation and are time consuming often resulting in a time lag between the point of sampling and the results being obtained. Therefore should the bioprocess not be proceeding optimally, this is often not discovered until after the window of opportunity for corrective action to be taken has passed. In addition, these techniques require the

removal of a physical sample from the bioprocess creating an opportunity for sterility to be compromised.³⁶

1.3.1.2 Infrared Spectroscopy

Near infrared spectroscopy provides a non-destructive, non-invasive method of obtaining near real time process measurements.⁷⁵ These key attributes have made it an attractive technique for monitoring of bioprocesses.

1.3.1.2.1 Analyte Selection

Within a bioprocess system there are a number of options as to where near infrared spectroscopic techniques could be applied. They could be utilised to monitor the concentrations of multiple key substrates and products in the process. Alternatively the composition of the medium and changes within can be monitored or physical features such as the biomass levels can also be determined using near infrared spectroscopy.

1.3.1.2.1.1 Monitoring of Substrates & Products

Electing to monitor the concentrations of major substrates or products during the bioprocess has been widely documented.^{18, 27, 28, 31, 40, 76} In bioprocesses one of the key targets for monitoring are the carbon and nitrogen sources.

Models constructed to determine glucose and glycerol concentrations in submerged culture processes have been reported by multiple authors and using a variety of organisms.^{32, 35, 71, 77} Electing to monitor substrate concentrations has been relatively successful with average errors of prediction for glucose ranging between 3% and 5% depending on the spectroscopic region utilised and sampling approach.^{35, 71} When monitoring glucose, in particular the natural conversion between forms (mutarotation), had a negative impact on the resulting model (Figure 1-6).⁵⁵ During the process, conversion between the two stereoisomers of glucose was observed until

an equilibrium position of approximately 37% α -glucose and 63% β -glucose was reached.⁷⁸ Whilst this equilibrium was being established the varying amounts of each form introduced an additional spectral variation that the system was trying to model. This resulted in an increased error of prediction for glucose, however removal of these spectra from the model improved the error of prediction associated with glucose.⁵⁵

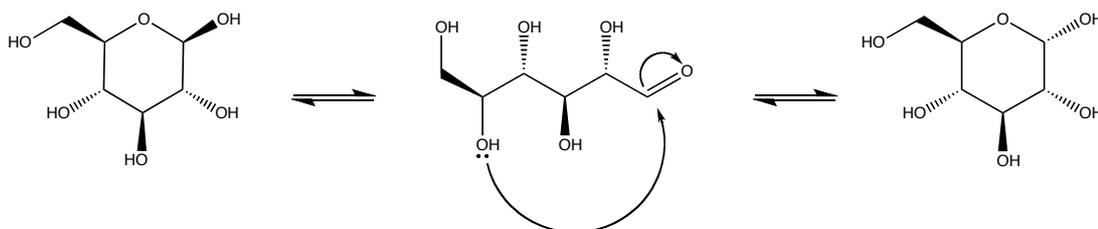


Figure 1-6 - Proposed mechanism for the conversion from α -glucose to β -glucose occurring in solution.⁷⁸

Other substrates such as ammonia concentrations in fermentation processes have also been monitored using this approach. These components have generally been present at much lower concentrations representing more of a challenge from a spectroscopic perspective. Despite this, with the aid of chemometric methods it has been possible to quantify even these low concentration analytes in complex biological systems.²⁴

Monitoring the products of a bioprocess using infrared spectroscopy has also been reported. Infrared spectroscopy has successfully been utilised to monitor the production of the biopolymer Gellan in good correlation with the analytical reference method.³¹ Also of particular note are the cases where ethanol has been successfully monitored using the technique. Since the OH vibrations are expected to be a key feature in any spectrum of ethanol, and this was also a dominating signal of the water background matrix, monitoring of this particular product was expected to be challenging. Despite this Mazarevica *et al.* (2004) and Sivakesava *et al.* (2001) have both reported successful models with reasonable associated errors.^{55, 62}

1.3.1.2.1.2 Monitoring of Culture Medium

Arguably the measurement of components in the bioprocess fluid could be regarded as the monitoring of substrates and products of the bioprocess. However when monitoring culture media multiple components in the system are generally measured, some of which are not present at the high concentration levels expected of a key substrate or product.

Reported applications of the technique in this manner have focused on quantifying components in mammalian cell culture fluids.^{49, 79} These systems require multiple essential components ranging from compounds similar to those required for bacterial cultures such as glucose, ammonia and lactate to amino acids such as glutamine.^{36, 49} The ability to measure the concentration of multiple components in the cell culture media is of great advantage in these processes. The multivariate nature of utilising near infrared spectroscopy also allows for monitoring of the cell culture process as a whole. Over the course of multiple replicates it was possible to identify a fingerprint for the process. This fingerprint could then be utilised along with the concentration data to implement a control strategy for the cell culture process.²⁸

Despite near infrared spectroscopy being able to successfully quantify multiple analytes within a system, the nature by which the signals arose and that they are broad and overlapping would be expected to limit the resolving power of the technique. Riley *et al.* (2001) suggested that independent quantification of a total of nineteen discrete amino acids based on their near infrared spectra was possible.³⁷ Quantification of these amino acids was achieved within an animal cell culture medium. Although these tend to be optically pure matrices, the restricted number of infrared active functional groups contained within amino acids and the broad overlapping nature of the signals would, from a theoretical perspective, question these results. Even using spectral pre-processing methods and multivariate calibration techniques (1.3.1.2.2 & 1.3.1.2.3) the limited variations in spectral features would likely result in some co-linearity within such a system. The methodology adopted by Riley *et al.* (2001) goes some way to addressing these

concerns with the use of synthetic medium samples rather than actual bioreactor samples. In doing so correlations in the concentration profiles could be minimised and the analytes monitored over a greater concentration range than would have been expected during a replicate of the bioprocess.³⁷

1.3.1.2.1.3 Monitoring of Physical Parameters

Spectroscopic measurements within the near infrared region have also been utilised in monitoring physical parameters of a biological process. Parameters such as biomass levels can be correlated with baseline shifts and offsets that result from an increased scattering of light by the cells, as well as distinct spectral features attributable to the cellular material.^{24, 25}

Monitoring of biomass levels has been successful and well documented adopting at-line and *in-situ* sampling approaches.^{24, 25, 30, 80} Arnold *et al.* (2002) discussed the construction of an at-line model for biomass and the associated challenges in then duplicating this model *in-situ*.²⁵ Despite these challenges an *in-situ* biomass model was constructed using spectra obtained from a transfectance probe. Errors of prediction returned by this model were comparable with those obtained from the at-line model.²⁵ The technique has also been successfully applied to a number of different bacterial cultures such as *Escherichia coli*, *Vibrio cholerae* and *Lactobacillus*.^{16, 23-25, 30, 40}

1.3.1.2.2 Spectral Pre-processing

The broad and overlapping nature of spectral features, particularly in the near infrared region, can be problematic from a modelling perspective. Spectral drift and the dominating signals resulting from water in the background matrix can obscure less pronounced but important spectral features.²⁴ These subtle spectral features can be enhanced and the baseline drift reduced by applying pre-processing algorithms to the spectroscopic data.

The most commonly employed pre-processing approaches are to smooth the spectra or calculate a derivative of the spectra and also apply the smoothing. Applying only smoothing can be useful and improve the spectra in cases where retaining the spectral drift is beneficial – such as in biomass models.⁸¹ In other cases however removal of the baseline drift is more beneficial, in which case the calculation of a derivative of the spectrum can be carried out. Derivatisation of the spectra fits a polynomial through the spectral data over a specified window size. Applying a Savitzky-Golay filter to the spectra data is the most common approach to smoothing and derivatising the spectra, although other approaches such as the Norris method can also be utilised.^{48, 81}

Other pre-processing approaches such as multiplicative scatter correction (MSC) and standard normal variate (SNV) can also be applied to enhance the obtained spectral data by attempting to deal with baseline shift in the spectra.⁸²

1.3.1.2.3 Calibration Strategy

Generally the calibration approach adopted has been to carry out multiple replicates of the bioprocess and correlate the spectroscopic measurements with values obtained from the reference analysis.¹⁷ A total of five process replicates was regarded as a sufficient number for the construction of a robust process model.⁸³

Adopting such a calibration procedure however risks introducing co-linearity into the model. Co-linearity occurs when the spectral features of one analyte are used to infer the concentration of another analyte. In biological systems such relationships are common, so it is particularly important to address the issue of co-linearity within any constructed model. In the case of a biotransformation system the concentration of reactants and products are stoichiometrically linked – as the concentration of reactant decreases the concentration of product increases in a 1:1 ratio. In this case a co-linear model may use the spectral features of the product to estimate the reactant concentration. Such a situation is undesirable as the model isn't dealing with each analyte independently, and if presented with a sample where the analytes are not

stoichiometrically linked may result in the incorrect quantification of one, or both of the analytes in question.

Some authors have reported the use of real, unaltered bioprocess data in the construction of a calibration model. Tamburini *et al.* (2003) put forward the case that these stoichiometric links in analyte concentrations were characteristic of the bioprocess and should be included in the model.⁴⁸ It was also postulated that given this was a biological process, concentrations of the various analytes would naturally vary. Arnold *et al.* (2003) acknowledged the importance of breaking co-linearity within the system however with an *in-situ* sampling approach retrospective adulteration of the samples was not an option nor was it consistent with the philosophy of not requiring sample preparation.^{17, 49}

Riley *et al.* (1997³⁶, 1998³⁸, 1998⁸⁴) have extensively investigated co-linearity within biological systems. Multiple methods of addressing this co-linearity within the system have been proposed with varying degrees of success.

Initially it was proposed that correlations within the system could be effectively dealt with by constructing a calibration model based on purely synthetic samples.^{38, 79, 84} This approach however was unsuccessful at predicting the analyte concentrations when presented with unseen real process samples. Failure to accurately predict these concentrations was not unsurprising since the variations and spectral features that arose due to components of the media or the biomass were not represented in the calibration model.

An adaptation to the purely synthetic approach that lay in the middle ground between synthetic samples and the real bioprocess samples was also proposed.^{38, 79, 84} This ‘spiking’, ‘semi-synthetic’ or ‘adaptive calibration’ approach involved the retrospective addition of varying concentrations of key analytes to actual bioprocess samples.¹⁷ Adopting this approach saw improvements to the errors of prediction of the associated model compared with a calibration based only on synthetic samples.

This approach, and variations on this approach, has been employed by other authors to achieve similar results.^{17, 21, 27}

The use of mathematically generated spectra has also been considered as a possible option for breaking co-linearity within these bioprocesses. In this approach spectra at various concentrations and mixtures, not necessarily at which an experimental spectrum had been obtained, are simulated based on spectroscopic exemplars.⁷⁹ With this approach there will always be the concern that the mathematically generated samples are not truly representative of the spectrum at this concentration, particularly when multiple analyte mixtures are under consideration.

When constructing a calibration model for any bioprocess there are a number of key recurring themes. The nature of these processes mean that stoichiometric linkages within the system do exist and must be adequately addressed to ensure the resulting model can independently predict the concentration of these analytes. In addition it is essential that the spectra included within the calibration model encompass the natural variation expected within the system. Ensuring the calibration model is sufficiently populated with spectra representative of the expected variation is essential for the construction of a reliable and robust calibration model.^{85, 86}

1.3.1.2.4 Near Infrared

The application of near infrared spectroscopy to a variety of challenging submerged culture processes has been extensively reviewed.^{19, 21} Process monitoring and control in this spectroscopic region has encompassed the three major sampling mechanisms (off-line/at-line, in-line and *in-situ*), considered a variety of organisms and analytes, as well as dealing with a number of important challenges. Some novel applications of the technique, challenges and solutions are discussed.

Macaloney *et al.* (1996) reported the application of at-line near infrared spectroscopy not only to monitor a submerged culture process but also to make process control decisions based on the results.¹⁶ During the bioprocess pH control was achieved by

the addition of ammonium hydroxide. However excess concentrations of ammonia within the system caused an inhibitory effect. Based on the at-line spectroscopic predictions the authors were able to identify inhibitory ammonia levels and switch to sodium hydroxide for pH control. Identification of a faulty pH sensor, one of the traditional mechanisms of process control, could also be detected based on the spectroscopic measurements. Similarly the inhibitory effects of acetate on growth could also be controlled. Determination of high acetate levels indicated a requirement to stop the glycerol feed forcing the cells to metabolise the acetate. When low acetate levels were predicted from the spectra glycerol feeding continued.

Arnold *et al.* (2000) reported the successful application of at-line near infrared in optically challenging matrices involving filamentous fungi.³² This particular fermentation process involved high water content, vigorous agitation and high aeration rates all common with submerged culture bioprocesses. In addition this was a multi-phase system with the presence of mycelia.³² Arnold *et al.* (2001) have also proposed an alternative and somewhat unique approach to bioprocess monitoring. In most cases the bioprocess has been treated as a single process from inoculation through growth and into the stationary phase. The authors proposed the construction of a separate model for each stage of the bioprocess. In doing so they reported an overall improvement in the resulting models for dealing with certain systems.

Adopting an in-line sampling approach Harthun *et al.* (1998) reported monitoring five key media components (glucose, lactate, ammonia & both D/L-glutamine) and the bioprocess product (antithrombin III) simultaneously using infrared spectroscopy. Predicted concentrations from the model were generally in good agreement with those determined by the off-line instrumental technique.²⁸

1.3.1.2.5 In-situ Near Infrared

Key process parameters such as biomass levels and the key media components of mammalian cell culture media can also be monitored using *in-situ* sampling approaches.^{23, 25, 49}

Tamburini *et al.* (2003) reported the successful application of an *in-situ* sampling approach to simultaneously monitor the biomass levels and concentrations of three additional media constituents (glucose, lactic acid & acetic acid) in near real time.⁴⁸ Of notable interest in this particular approach was the use of no spectral pre-processing techniques to enhance spectral features or reduce the dominance of water. In doing so the authors were not only successful in quantifying the key components of interest but, due to the simplicity of the approach, were able to do so in near real-time due to a lack of complex and time consuming mathematical manipulations.

1.3.1.2.5.1 Multiplexing

In-situ near infrared spectroscopy allows for multiplexing, a process where a single spectrometer system is utilised with many reactor systems.⁴⁶ On the industrial scale this represents substantial cost savings over the dedication of one spectrometer system per reactor. The setup essentially shares the detector between multiple reactors, with each system connected to the spectrometer via its own dedicated fibre optic bundle and *in-situ* probe (Figure 1-7).^{42, 46}

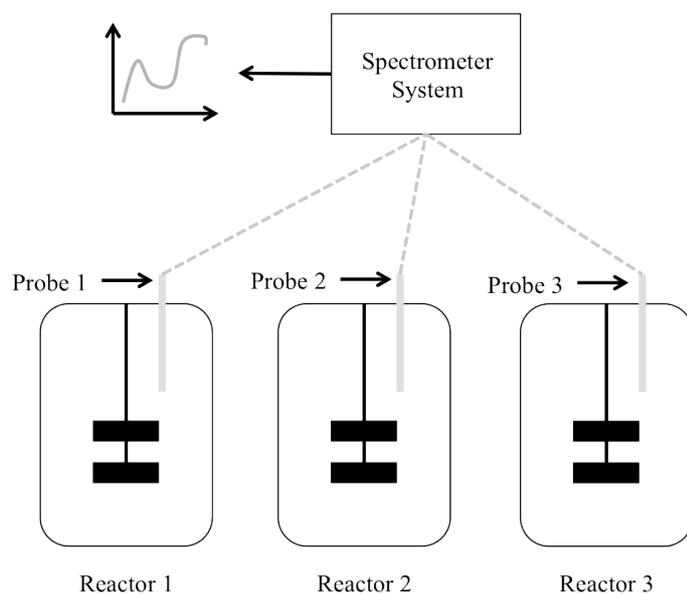


Figure 1-7 – Schematic representation of multiplexing setup adapted from Chen *et al.* (2011).⁸⁷ One spectrometer system is used to acquire the infrared spectra from three

different reactor systems. Spectra can then be used with previously constructed multivariate models for the quantification of multiple analytes in each reactor.

Operating the spectrometer in this multiplexed fashion does not affect the ability of the system to quantify multiple analytes within the system simultaneously. The infrared spectrum acquired can still be utilised with multivariate calibration approaches to simultaneously predict many analytes in the same manner as a dedicated system can be utilised. The multiplexing approach merely offers a cost and efficiency benefit over dedicating a spectrometer system to each reactor system.

As well as cost efficiencies multiplexing offers a rapid means of populating a calibration model, allowing many process replicates to be carried out at the one time making efficient use of the available instrumentation. There are however a number of drawbacks with this approach.

The largest source of variation in the system is more likely to be due to optical differences between sampling probes rather than the intended process variations. Roychoudhury *et al.* (2007) illustrated the optical differences of seven different *in-situ* probes in simple chemical solutions of acetone and methanol.⁴⁶ To compensate these variations were built into the resulting model, which returned reasonable errors of prediction for the unseen datasets.

Instrument variability and calibration model maintenance have been well-documented issues in both the near and mid infrared regions. However many of the noted difficulties can be overcome by carrying out instrument standardisation. This approach suitably dealt with variations not only between probes but also between spectrometer systems. In this respect a model can be rapidly updated following instrument maintenance or replacement without requiring reconstruction of the model.⁸⁸

1.3.1.2.6 Mid Infrared

Within the mid infrared region the tendency has been more towards monitoring of key analytes within the system such as substrate, products or medium components.

Monitoring of physical process parameters have also been reported using mid infrared spectroscopy. Parameters such as biomass levels have typically not been monitored using this approach due to the short path lengths required to obtain suitable spectra.⁶⁴ Schenk *et al.* (2007) have however reported making pH estimations based on the relative absorbance of the protonated and deprotonated forms of the acid.⁸⁹

Unlike the near infrared region where it was unusual for the spectra to be utilised without any spectral pre-processing techniques, the mid infrared spectra can be utilised directly in the model. However depending on the process conditions enhancement of the model can be achieved by derivatisation. Models constructed from the unmodified spectra were able to accurately predict the substrate and product at-line in a *Pichia pastoris* submerged culture.⁵³

Similar results have been obtained in systems employing a variety of organisms such as *Escherichia coli*, *Saccharomyces cerevisiae* and *Lactobacillus casei*.^{56, 60, 61}

An example of process control using mid infrared spectroscopy was reported by Schenk *et al.* (2007). A diamond ATR probe was housed within a flow through cell and the collected mid infrared spectra used to calculate the methanol content of the reactor. These values were then transferred to a control system, which altered the methanol feed rate to maintain a constant concentration within the system.⁶⁴

Application of mid infrared spectroscopy *in-situ* has been reported, however these have mainly focused on proving the techniques ability to quantify multiple key components of the system rather than process control.^{56, 68} The challenges in developing fibres able to transmit light in this region may offer some explanation as

to why the technique has not yet been widely implemented as a process monitoring and control strategy.

1.3.2 Biotransformation Processes

Most of the applications of infrared spectroscopy discussed have been in relation to submerged culture processes. All the sampling mechanisms discussed previously and many of the challenges and associated difficulties reported will be applicable to these biotransformation systems. Additionally the subtle molecular differences and stereochemistry were also likely to present a number of challenges both in terms of spectroscopy and modelling that were not observed with the other bioprocesses discussed.

1.3.2.1 Near Infrared Spectroscopy

Despite near infrared spectroscopy being a very popular technique for monitoring submerged culture processes its application to biotransformation systems has been limited. Presumably this is a result of the broad, overlapping and unspecific nature by which signals in this region arise.

Bird *et al.* (2002) reported the application of near infrared spectroscopy to monitor a Baeyer-Villiger type biotransformation process. This particular process also illustrated the ease with which microbial systems were able to achieve such conversions compared with synthetic organic chemistry routes.

Synthetically the Baeyer-Villiger reaction results in the formation of an ester linkage via insertion of an oxygen atom by reaction of a ketone and peracid (Figure 1-8). The reaction proceeds via a proposed alkyl migration step, with the migrating group determined based on ability to stabilise the positive charge. Stereochemical control in Baeyer-Villiger type reactions tends to favour retention of the substrates stereochemistry, even in cases where alteration of stereochemistry would result in a more energetically favourable structure.⁹⁰ Generation of product esters with a

specific stereochemistry can be obtained by utilising transition metal catalysts during the synthetic process.⁹¹

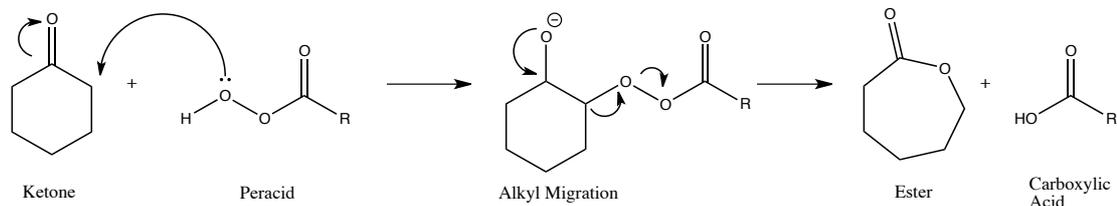


Figure 1-8 - Mechanism of typical Baeyer-Villiger reaction carried out chemically.

Microbial conversion of an asymmetric ketone (specifically bicyclo[3,2,0]hept-2-en-6-one) using Baeyer-Villiger monooxygenases was reported by Doig *et al.* (2003) (Figure 1-9).⁹² In this case however two resulting lactone products were formed, regioisomers of each other. It was also noted that each regioisomer was formed as a single stereoisomer, a phenomenon that was not possible to replicate using conventional synthetic approaches.⁹²

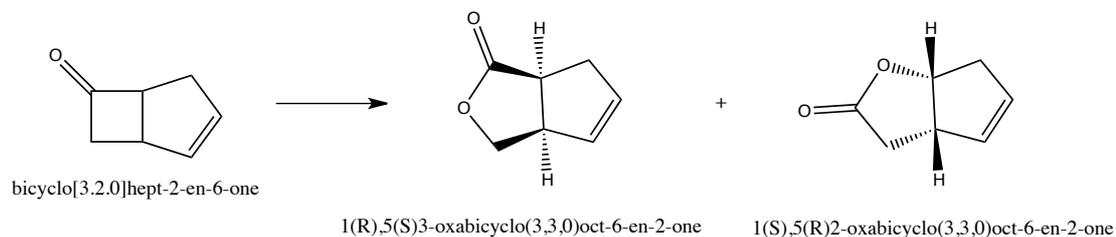


Figure 1-9 – Conversion of ketone to lactone using microbial Baeyer-Villiger monooxidase enzyme as described by Doig *et al.* (2003) and monitored spectroscopically by Bird *et al.* (2002).^{92, 93}

Bird *et al.* (2002) reported monitoring of this particular biotransformation process using near infrared spectroscopy. Acquisition of pure component near infrared spectra of the two analytes of interest suggested that the majority of spectral features observed were common to both compounds.⁹³ Given the structural similarities between the two compounds this was not unexpected, however some minor variations were observed. To enhance the spectral features the acquired spectra were

converted to the second derivative format. Other approaches, such as referencing the spectra to a water background as opposed to air, were considered but rejected based on background stability.⁹³ Resulting models constructed from both at-line sampling and *in-situ* sampling with a transfectance probe were reported. Despite the similarities in the spectra the resulting models were able to predict the concentration of both compounds with errors of prediction in good agreement with the errors of calibration.⁹³

1.3.2.2 Mid Infrared Spectroscopy

The use of mid infrared spectroscopy has been more commonly reported for monitoring biotransformation processes. Within this region the spectra obtained could be used as a chemical fingerprint for the analytes present in the system.⁹⁴

Key analyte concentrations in a relatively simple biotransformation process have been monitored using *in-situ* mid infrared spectroscopy. Degradation of the cyanide-containing compound by *Rhodococcus rhodochrous* can be achieved using two different routes by the microorganism (Figure 1-10). Firstly a direct route converts the cyanide to the corresponding carboxylic acid and ammonia using the nitrilase enzyme. Alternatively a two-step process can be adopted where nitrile hydratase converts the cyanide to an amide. This is then followed by conversion to the carboxylic acid and ammonia by amidase enzymes.⁹⁵

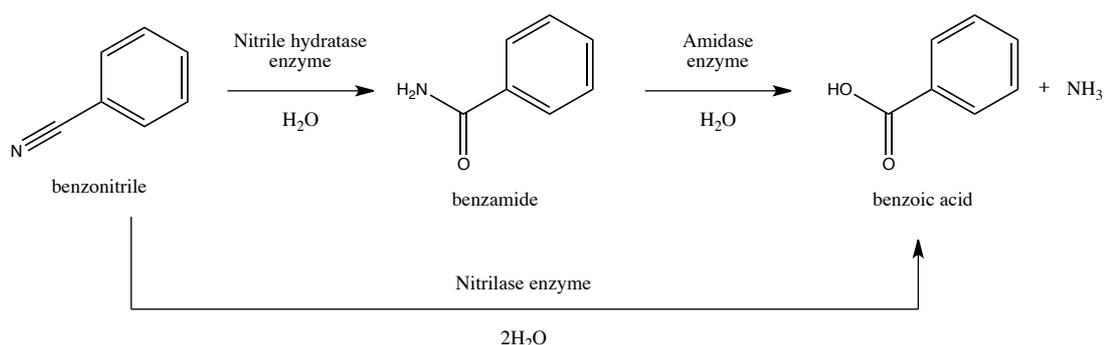


Figure 1-10 - Routes for the degradation of cyanide containing compounds to the corresponding amine and carboxylic acid by *Rhodococcus rhodochrous* reported by Dadd *et al.* (2000).⁹⁵

Over the course of the biotransformation process changes in spectral features were clearly identified. Signals attributable to the benzoic acid were noted to increase over the time course of the bioprocess. Features due to benzamide were observed to increase followed by a decrease in intensity as the process progressed, which was consistent with degradation of the cyanide via the two step process. At this stage however the process was only considered qualitatively, there were no quantitative models for the biotransformation process constructed.⁹⁵

In a slightly more complex biotransformation process quantitative models for the determination of 1-phenyl-1,2-propanedione substrate and product (1R2S)-1-phenyl-1,2-propanediol were constructed (Figure 1-11). Generation of the product with this particular stereochemistry was required as an intermediate in the synthesis of the stimulant ephedrine.⁷²

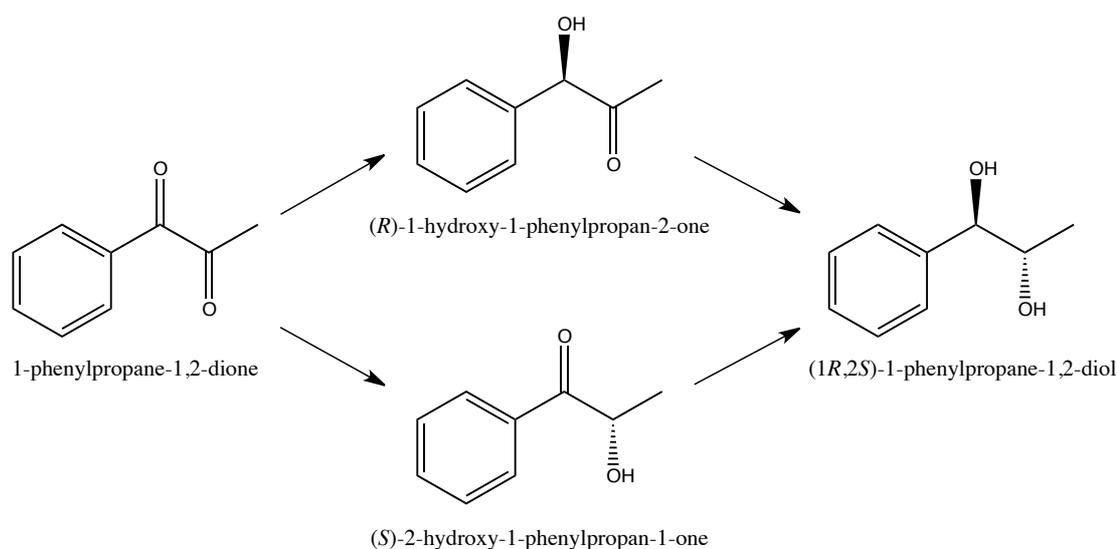


Figure 1-11 – Biotransformation process for the conversion of 1-phenylpropane-1,2-dione to (1R,2S)-1-phenylpropane-1,2-diol used in the synthesis of the stimulant ephedrine.⁷²

The biotransformation process was carried out using a whole cell suspension of *Saccharomyces cerevisiae* with the spectra collected using an ATR crystal located within a flow through cell. Since this was a quantitative model reference measurements to determine the concentrations of substrate, product and intermediate

compounds were determined by gas chromatography mass spectrometry (GC-MS). Since the spectra were being acquired continually the intermediate analyte concentrations were determined by interpolation using a non-linear fit.⁷²

Quantification of the substrate and product based on the spectroscopic data was achieved with good agreement between the predicted and measured concentrations.⁷² These compounds were structurally quite different with the product containing the strongly absorbing hydroxyl functional group not observed in the substrate. The constructed model also attempted to quantify the concentrations of the intermediate materials. Prediction of these intermediate compounds was not as good however, this was not surprising given the low concentrations at which these were expected and that spectral variations between the two were likely to be limited. In this particular case, the limited number of samples removed and the absence of any steps to deal with co-linearity within the system suggested possible inference of these concentrations as opposed to actual measurements.

Difficulties in distinguishing between compounds where the molecules differ at a single functional group or in their stereochemistry using infrared spectroscopy was likely to be challenging. Macauley-Patrick *et al.* (2003) reported the application of at-line mid infrared spectroscopy to monitor D-sorbitol and L-sorbose concentrations during a *Gluconobacter suboxydans* biotransformation process. These compounds differed at two positions on the molecule: by functional group (carbonyl and alcohol) at the C₂ position and stereochemistry at the C₅ position. Examination of the pure component confirmed unique spectral features could be identified for both analytes of interest. In conjunction with multivariate modelling approaches a suitable model for the system was constructed.⁹⁶

Application of this technique to biotransformation processes has highlighted the power of mid infrared spectroscopy for monitoring these processes where the compounds of interest are only subtly different, and in some cases are present at low concentrations. Widespread application of infrared spectroscopy to this particular classification of bioprocess is still within its infancy. However the potential for

applying the technique to other complex, optically challenging, environmentally inhospitable and industrially relevant biotransformation processes is already evident.

1.3.3 Alternative Process Monitoring Approaches

A number of alternative process monitoring techniques have been proposed. Many of these however are still based around the interaction properties of light with the sample material.

Fluorometry or fluorescence spectroscopy has been reportedly employed to monitor bioprocesses.⁹⁷ In order for this approach to be successful the analytes of interest must exhibit fluorescence properties, meaning there must be specific absorbance and emission frequencies. Within biological systems components such as amino acids, proteins and coenzymes such as ATP all exhibit fluorescence properties. This can be advantageous if these compounds are the target analytes or challenging if these materials are exhibiting interference effects.⁹⁷

Multiple components within systems can be monitored by utilisation of two-dimensional fluorescence spectroscopy. Development of an *in-situ* sampling approach has also been reported.⁹⁸ Kara *et al.* (2010) have also reported application of this technique to biotransformation processes and the construction of quantitative models for the system.⁹⁹

Raman spectroscopy, a technique complementary to infrared spectroscopy, has also been utilised for monitoring and control of bioprocesses. Biological matrices pose particular challenges in the application of Raman spectroscopy not only because of light scattering caused by the cellular material and dissolved gases but by the fluorescent nature of the sample which can obscure the Raman scatter. Despite these challenges however quantitative models of bioprocesses utilising this technique have been reported.^{97, 98}

Microscopy has also been developed and applied *in-situ* to obtain images during the bioprocess. The approach has been useful in the determination of biomass levels, cell counts and the identification of morphological changes within the system.^{97, 98}

Refractive index measurements can also be utilised to monitor critical process parameters. The measurement approach was substantially different from that employed by bulk refractive index detector systems. The system was based on a tapered fibre through which the laser light source was directed. Detector systems either end of the tapered fibre identified changes in the transmitted light through the fibre, as a result of the cellular material attaching to the fibre, which were utilised to calculate a refractive index for the system. This particular monitoring approach has been reported as an alternative method for determining biomass levels during a bioprocess.¹⁰⁰

Moving away from light based measurement approaches Maskow *et al.* (2006) reported the application of calorimetry to monitor biotransformation processes. Using a miniature heat sensor connected to the bioreactor via a flow through cell measurements of heat evolution were made by comparison against a stable reference value. Various parameters were then calculated based on the kinetics of the system.¹⁰¹

1.4 Chiral Amino Acids

The biotransformation processes under investigation in the current study deal with the generation of chiral amino acids. Amino acids are essential molecules often regarded as ‘building blocks’ due to their role in the production of proteins.¹⁰² Due to commercial sensitivity the use and application of the amino acids investigated has not been disclosed. Based on information obtained in the literature, the potential applications of industrially generated amino acids are discussed suggesting potential applications of the generated amino acids.

Amino acids with a single stereochemistry have a number of applications across a variety of industries. In the food industry chiral amino acids are utilised to generate synthetic sweeteners as well as flavourings (amino acids react with sugars to produce compounds that mimic spices).^{102, 103} These amino acids are also important compounds for the production of probiotic cultures for use in foodstuffs.¹⁰²

The cosmeceutical industry has also expressed interest in the generation of chiral amino acids, specifically the production of L-cysteine for use in anti-aging preparations.¹⁰²

Another industry where the large-scale production of chiral compounds is of great interest, and importance, is the pharmaceutical industry. D-enantiomers of amino acids have been produced microbially for the production of synthetic derivatives of penicillin. The microbial production of these compounds can be achieved either using de-racemisation enzymes to convert the unwanted enantiomer of a racemic starting material, or using racemase enzymes that actually flip the stereocentre of the, in this case, undesired L-configuration.^{102, 103}

Natural and unnatural L-amino acids for pharmaceutical applications can also be prepared in a similar manner. Preparation and isolation of the L-enantiomer tends to be performed by de-racemisation enzymes acting on racemic substrate.¹⁰² This is most likely a result of the cells ability to utilise the L-enantiomer directly, thereby having no need to convert the D-form.¹⁰⁴

Although the use of the generated chiral amino acids has not been disclosed, the above discussion gives some indication as to the potential industrial requirements to generate amino acids with a given stereochemistry.

1.5 Aims & Objectives

Firstly the critical review of the literature carried out was utilised to determine the current extent to which infrared spectroscopic techniques were utilised for

monitoring and control of bioprocesses. Many of these processes were submerged culture systems with limited application to biotransformation processes, however these models tended to be either qualitative or the substrates and products of the biotransformation process fairly distinct.

The aim of the current research was to investigate the potential of applying both near and mid infrared spectroscopic techniques to monitor and control various industrially relevant amino acid de-racemisation biotransformation processes. Using the de-racemisation process described by Fotheringham *et al.* (2006) the production of various un-natural amino acids was considered. Various spectral acquisition approaches were investigated in an attempt to construct and validate reliable, robust models for the system thereby introducing a QbD approach to this industrial biotransformation process.

Ultimately for each biotransformation process investigated it was intended to construct a robust calibration model for the process. This model would be challenged using unseen process replicates in order to assess the predictive ability of the model. It was also desirable to further challenge the model by presenting process replicates carried out off-site and assessing how well the predicted concentrations correlated with those measured using off-line reference analysis.

2. Background & Theory

2.1 D-amino Acid Oxidases

D-amino acid oxidase (DAAO) enzymes are stereospecific enzymes that act upon the 'D' enantiomer of amino acids, converting the amine functional grouping to a ketone thus generating the corresponding α -keto acid. These enzymes do not exhibit any activity towards the 'L' enantiomer of amino acids, however separate L-amino acid enzymes have been isolated and exploited commercially.^{105, 106}

Expression and overproduction of the enzyme has been carried out in *E. coli* and yeast cells by the modification and insertion of a plasmid containing the appropriate sequence for production of the enzyme.^{4, 106} These cells can then be cultured to a suitable biomass level and induced to trigger over-production of the desired DAAO enzyme.

During the biotransformation phase the D-amino acid for conversion enters the cells where the biotransformation process can be undertaken. Conversion of the D-amino acid to the corresponding α -keto acid proceeds via the α -imino acid in the proposed mechanism (Figure 2-1). The amino acid is oxidised to the corresponding α -imino acid with the assistance of flavin-adenin dinucleotide (FAD), which is itself reduced in the process. Reaction with water will convert the formed α -imino acid into the corresponding α -keto acid product, which will then be exported from the cell. The reduced FAD undergoes oxidation with molecular oxygen to replenish the FAD with hydrogen peroxide resulting as a by-product. Catalase, if present, breaks down this formed hydrogen peroxide to yield water and oxygen.^{3, 105}

Over expression and utilisation of the DAAO enzyme has been exploited for the production of α -keto acids as well as for de-racemisation processes for the generation of chiral compounds. However, process economics have a substantial influence on the suitability of this approach.^{105, 107}

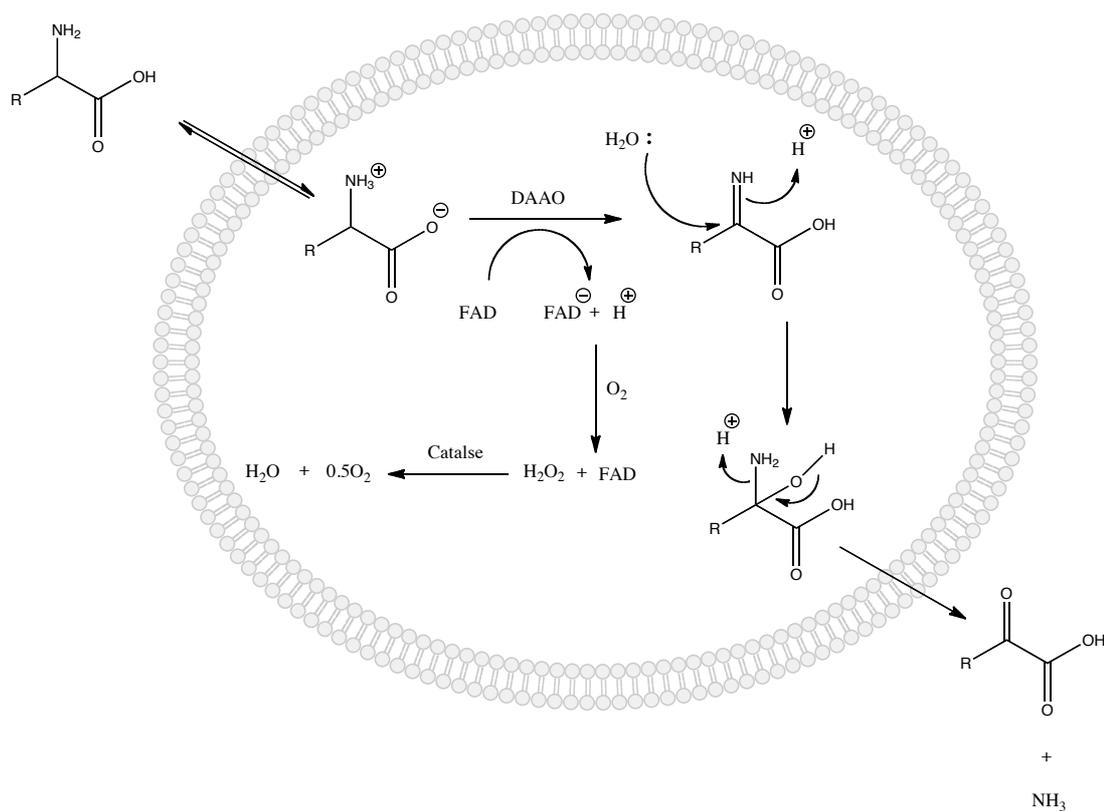


Figure 2-1 – Proposed mechanism for the conversion of D-amino acid substrate into the corresponding α -keto acid proceeding via an α -imino acid intermediate and catalysed by DAAO enzyme and FAD adapted from Garcia-Garcia *et al.* (2008)¹⁰⁵.

2.2 Transaminase Enzymes

Microorganisms employ transaminase enzymes for the generation of amino acids for use in the synthesis of proteins. These enzymes do not have a high degree of specificity, and have been reported to work successfully on amino acids with a range of side groups.^{108, 109}

The amine functional group from the donor amino acid is transferred to the substrate keto acid, with the retention of stereochemistry, forming the desired stereospecific amino acid compound. In the process the donor amino acid is itself converted to the corresponding keto acid.

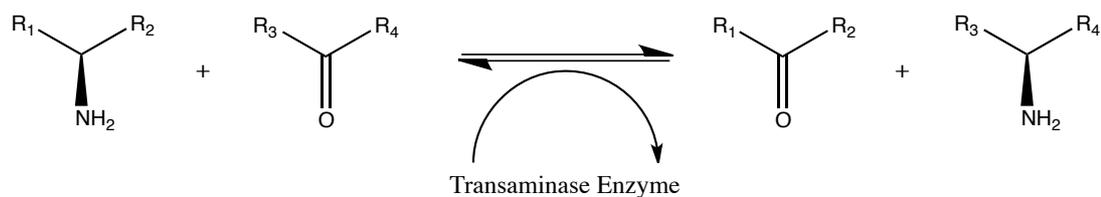


Figure 2-2 - Transaminase biotransformation process illustrating conversion of substrate keto acid to the desired amino acid with retention of stereochemistry.

2.3 High Performance Liquid Chromatography

Chromatography is an instrumental technique that allows for the separation, detection and quantification of a sample mixture into its component parts. In the liquid phase separation is achieved by establishing a partitioning effect between the stationary and mobile phases.

The various components of a HPLC system are represented schematically in Figure 2-3.

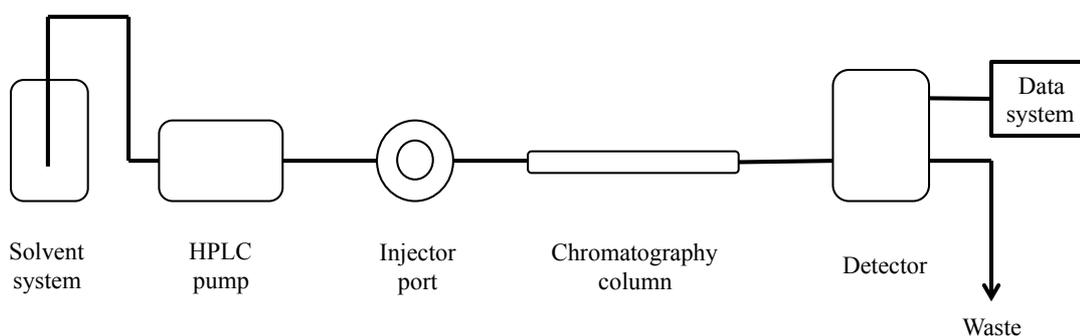


Figure 2-3 - Schematic representation of the main components in a HPLC system adapted from Hamilton *et al.* (1982)¹¹⁰.

Mobile phase (eluent) is stored within solvent reservoirs, depending on the set-up of the system this can range from a single solvent for isocratic methods through to multiple solvents utilised in gradient elution methods. In all cases the solvents utilised should be of high purity and solvents de-gassed prior to use.

Eluent is transported through the chromatography system using a pump, or series of pumps. The pumps utilised are designed to ensure a consistent and pulse free flow of mobile phase through the chromatography column.¹¹⁰

Samples are introduced into the system via a six-port syringe-loop (Rheodyne) design valve. Under normal conditions (valve in the load position) the eluent flows through the port and onto the column, with the injector aligned so that any sample injected passes through the loop and out to waste, allowing the sample to be loaded without disrupting the eluent flow through the column. The sample loop has a fixed volume and ensures that a consistent sample volume is introduced into the system with each injection. When rotated to the inject position the eluent flow is rapidly diverted to pass through the sample loop, taking the contents onto the chromatography column for separation.¹¹⁰

Separation of the various analytes is achieved in the chromatography column. The commonly utilised reverse phased column is packed with small particles of silica that have had the hydroxyl group end-capped by reaction with a C₁₈ alkyl chain. As analytes pass through the column, they are separated based on their varying distribution and affinity between the stationary and mobile phases.

Various detector systems have been employed with liquid chromatography systems, and fall into two categories; detector systems that measure a specific property, such as diode array detectors (DAD), and systems that measure a bulk change in the eluent properties, such as refractive index detector systems.¹¹⁰

2.4 Vibrational Spectroscopy

Spectroscopy in general can be described as the interaction of light with a sample.¹¹¹ Within the electromagnetic spectrum frequencies of light in both the visible and infrared regions are commonly utilised for spectroscopic purposes. The infrared region can be further split into three distinct regions: the near infrared (4,000 cm⁻¹ through 14,285 cm⁻¹), mid infrared (400 cm⁻¹ through 4000 cm⁻¹) and far infrared

region (4 cm^{-1} through to 400 cm^{-1}). Both the near and mid infrared regions are commonly utilised for spectroscopic applications.¹¹¹

The vibration of bonds within a molecule can be considered at a superficial level or at a more detailed level where the energy levels of the molecule and excitations between these energy levels are considered.

In a simplistic approach the exposure of a molecule to radiation can bring about a movement of the bonds between the atoms of that molecule. The bonds can move in a number of different ways, they can rotate, vibrate or move along a plane. Light in the infrared region of the spectrum causes the bonds within a molecule to vibrate, hence the technique is often referred to as vibrational spectroscopy. The total number of degrees of freedom (possible movements) associated with a molecule is three times the number of atoms within the structure. Of these possible movements three are translational motion and a further three are rotational motion both along the X, Y and Z axis. The remainder of the molecules possible degrees of freedom are due to vibrational motion. Therefore, the total number of potential vibrations exhibited by a molecule can be determined using Equation 5. In cases where the molecule is linear there is no rotation about the axis in which the bond lies therefore the total number of vibrations is given by Equation 6.¹¹²

$$\text{Number of Vibrations} = (3N - 6) \qquad \text{Equation 5}$$

$$\text{Number of Vibrations} = (3N - 5) \qquad \text{Equation 6}$$

Vibration of the bonds between two atoms within a molecule can occur via a number of possible modes. Two stretching modes can occur: symmetric and asymmetric. Four possible deformation modes exist: scissoring; rocking; twisting and wagging (Figure 2-4).^{113, 114} In order for these vibrations to be active and result in a signal in the infrared spectrum there must be a change in the molecules dipole moment (μ) as a result of the vibration.^{112, 115}

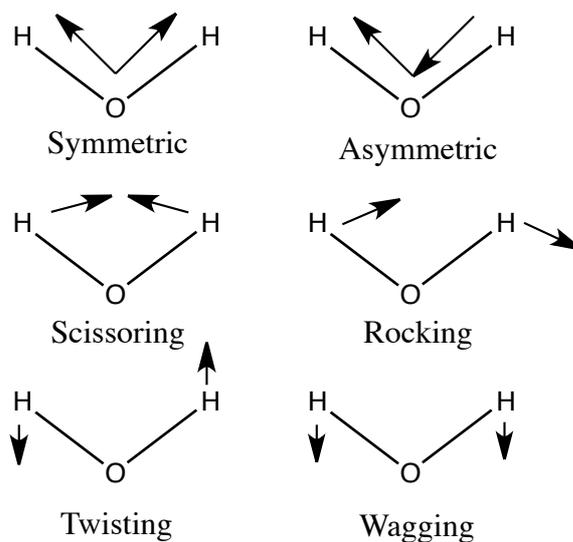


Figure 2-4 - Illustration of possible vibrations of bonds within a molecule adapted from Kemp (1987)¹¹³.

Considering a more detail approach, the energy levels that exist can be considered. There are three distinct energy levels that can be identified – the electronic, vibrational and rotational levels (Figure 2-5). The absorption of light equivalent to the difference between two of these energy levels can cause excitation from the ground state to one of these excited states. In the infrared region these excitations occur between the vibrational energy levels.⁵⁰

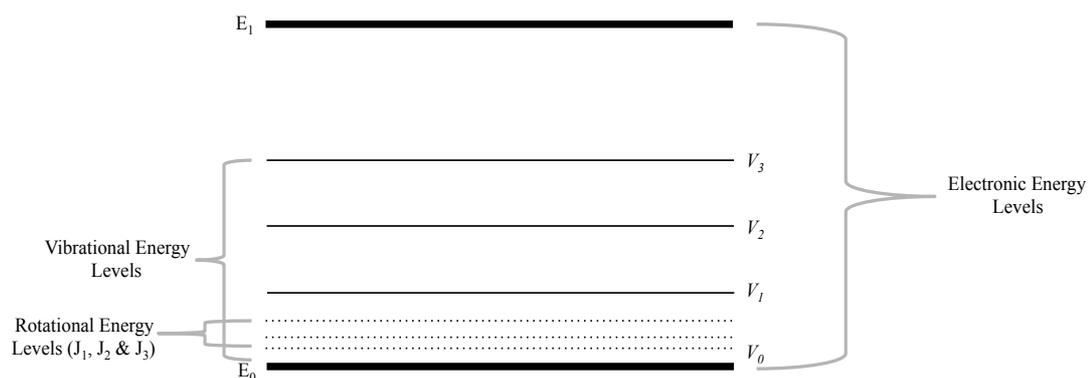


Figure 2-5- Schematic representation of the electronic, vibrational and rotational energy levels adapted from Crabb *et al.* (1995).⁵⁰

If a simple diatomic molecule is considered, the bond between the two atoms can be treated like a spring. As the bond length is increased or decreased there will be a change in the potential energy of the system. When a force is no longer applied the spring (or bond) returns to its equilibrium position with the energy released causing the bond to vibrate. This approximation assumes that the diatomic molecule is behaving as an ideal system (harmonic oscillator). Plotting the potential energy of the system against the bond length will generate a parabolic shape with the minima at the equilibrium bond length and the vibrational energy levels equally spaced (Figure 2-6).¹¹⁶ In this simple case transitions between the vibrational energy levels occur only by a single energy gap at a time ($\Delta v_i = \pm 1$) and by the absorption of radiation equal to the energy difference between levels.

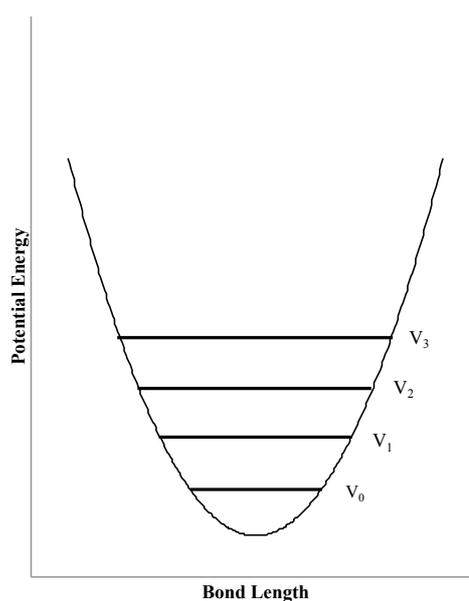


Figure 2-6 – Plot of potential energy against bond length for a simple diatomic molecule behaving as a harmonic oscillator adapted from Hollas (2002).¹¹⁶

In reality however even a simple diatomic molecule does not behave as an ideal system. Stretching a bond will eventually reach a point where the bond will dissociate, so the system must be treated as an anharmonic oscillator. In this case the plot of potential energy against bond length will eventually plateau at the point where dissociation occurs (Figure 2-7). The vibrational energy levels are also no

longer equally spaced but move closer together as the potential energy in the system increases.¹¹⁶ These deviations from the ideal situation mean that transitions between more than one vibrational energy level ($\Delta v_i = 2,3$) are possible (overtones) as well as the simultaneous vibration of multiple bonds by a single photon (combination bands) are now able to occur.¹¹²

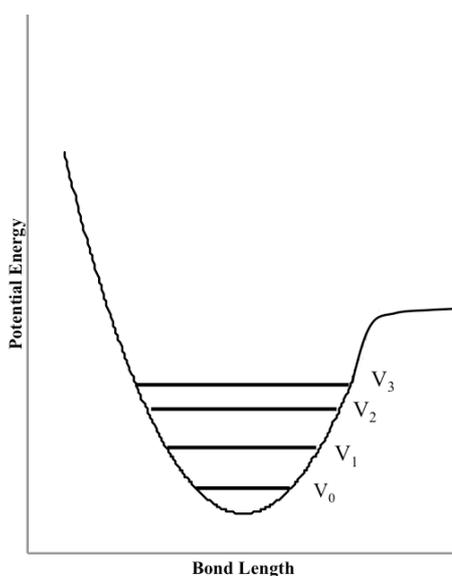


Figure 2-7 – Plot of potential energy against bond length for simple diatomic molecule behaving as an anharmonic oscillator. Energy levels are no longer equally distributed and bond dissociation is taken into consideration.

With the exception of enantiomers, every molecule has different vibrational modes which result in a unique spectrum that can be used for identification, structural elucidation and quantitative purposes.¹¹²

The process of spectroscopy is concerned with measuring the interaction of an incident light (radiation) beam with a sample material. Conventionally the resultant output spectrum is displayed as a plot of absorbance plotted against wavenumber, however the instrumentation actually measures the transmitted (unabsorbed) light and converts this to absorbance using Equation 7, where I_0 represents the intensity of the incident light and I the intensity of transmitted light.¹¹¹

$$\text{Absorbance} = \log_{10} \left(\frac{I_0}{I} \right)$$

Equation 7

The electronic effects associated with the absorption of light in the middle and near infrared regions of the electromagnetic spectrum and the resulting spectra are considered.

2.4.1 Middle Infrared

Within the middle infrared region (400 cm^{-1} through 4000 cm^{-1}) signals arise due to the fundamental vibrational transitions. Excitation is from the ground state to the first excited state ($\Delta v_i = \pm 1$) accompanied by the absorption of light equal in energy to the difference between these two states (Figure 2-8). These allowed transitions result in high intensity absorptions meaning only small sample thicknesses (approximately $10 \mu\text{m}$) are required to obtain a spectrum that is neither saturated nor excessively weak.¹¹²

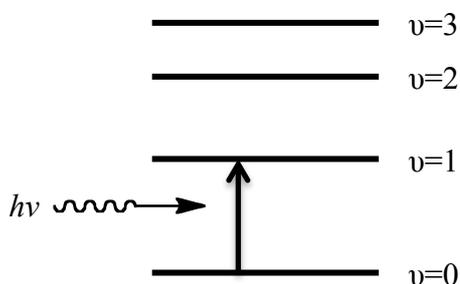


Figure 2-8 - Diagrammatic representation of vibrational transitions on absorption of appropriate light at frequency in the mid infrared region.

2.4.2 Near Infrared

Within the near infrared region ($4,000\text{ cm}^{-1}$ through $14,000\text{ cm}^{-1}$) the signals are due to overtones and combinations that arise due to the system behaving as an anharmonic oscillator. Absorption of radiation with the appropriate energy can result in transitions from the ground state to an excited state higher than the first ($\Delta v_i \neq \pm 1$). This phenomenon is what is observed as an overtones signal (Figure 2-9).¹¹²

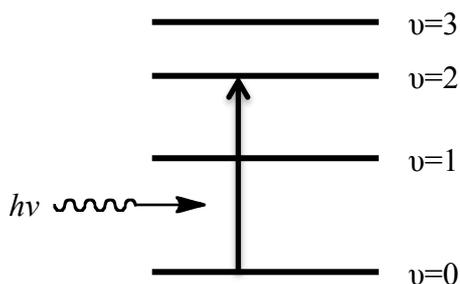


Figure 2-9 - Overtone transition from the ground energy level to the second excited state.

Spectral features as a result of combination bands are generally observed in the region between 4000 cm^{-1} and approximately 5100 cm^{-1} . Like the overtones, these features arise as a result of the molecules behaving as anharmonic oscillators. These signals are due to simultaneous vibrations that result from the absorption of light at an energy equivalent to a linear combination of the required energy gap (Equation 8).^{50, 114}

$$E = \nu_{(E_{0a}E_{1a})} + \nu_{(E_{0b}E_{1b})} \quad \text{Equation 8}$$

Both the overtones and combination bands signals are formally disallowed processes, but occur due to non-ideal behaviour of the molecule. As a result although these processes can happen they do not occur as frequently as the allowed transitions, such as those observed in the mid infrared region. Resulting signals in the near infrared

region are therefore weaker than the corresponding fundamental transitions allowing for longer sample path lengths to be utilised.¹⁹

2.4.3 Instrumentation

Infrared spectrometers are commonly available in two variations: Fourier-transform infrared systems (FT-IR) and dispersive infrared spectrometers.²¹ Near infrared spectrometer systems can be based around either the dispersive or Fourier transform instrument designs. However, instruments based on the Fourier transform design dominate the mid infrared instruments.

The same output spectra results from both system designs however, the mechanism by which this is achieved differs between the instrumentation. Both designs have their associated advantages and disadvantages.

2.4.3.1 Dispersive Instruments

Dispersive infrared spectrometers consist of three basic components (the radiation source, a monochromator and the detector system) arranged as illustrated in Figure 2-10. Instruments based on this design tend to be more common for the near infrared region.

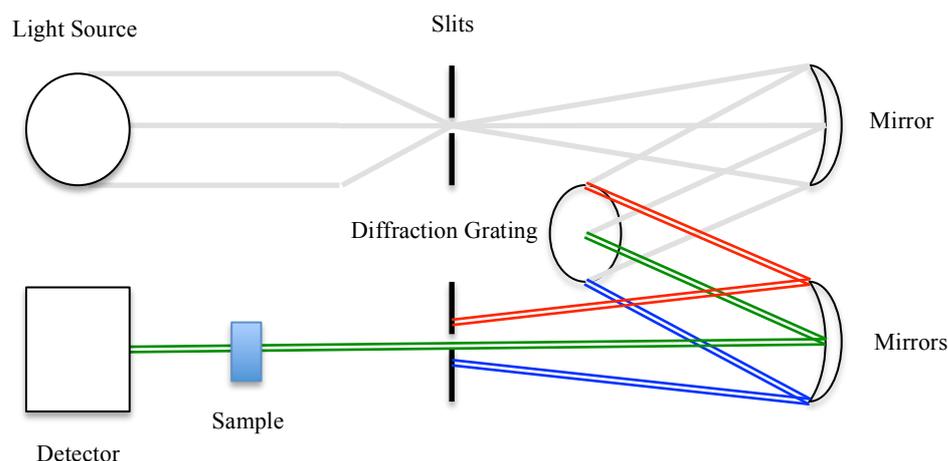


Figure 2-10 - Schematic representation of basic layout of dispersive infrared spectrometer and the component parts of monochromator unit illustrating how incident beam of light travels through the spectrometer system.

Due to the nature by which the infrared signals arise (2.4.2) the light source selected should provide continuous emission of all the desired frequencies within the identified region of interest, in this case the near infrared. This requirement can be achieved using a laser light source, or using sources such as the Zirconium point lamp.¹¹⁷

The monochromator unit is made up from a series of individual components with the function to take this continually emitted light and accurately split it into the individual frequencies before passing the light through the sample. The slits are used to control the intensity of incident light passed to the monochromator. Light is then directed towards a mirror that ensures the beams are parallel before directing them towards a diffraction grating. The diffraction grating splits the light beam up into the various frequencies, which are directed to the exit slit via another mirror. Variation of the light frequency at the exit slit can be achieved by moving the position of the diffraction grating.¹¹⁷

Dispersive spectrometer systems are only able to measure a small section of the infrared spectrum at any given time. Movement of the diffraction grating allows different frequencies of light to be selected thereby allowing the full spectral range to

be covered. This represents one of the main disadvantages of this system design in that movement can cause vibrations and temperature variations causing an effect on the acquired spectrum.²¹

2.4.3.2 Fourier Transform Instruments

Fourier transform instruments have the same basic layout as the dispersive instruments except the monochromator set-up is replaced by an interferometer (Figure 2-11). This consists of a beam splitter and a series of mirrors, some of which are fixed and others that move.

Incident light is directed to the beam splitter, which allows some of this light to be transmitted but is also able to reflect some of the incident light. The transmitted light is reflected off a fixed mirror and directed back towards the beam splitter. Non-transmitted light is reflected off a moving mirror before being directed back to the beam splitter where the light beams are combined (or exhibit interference).

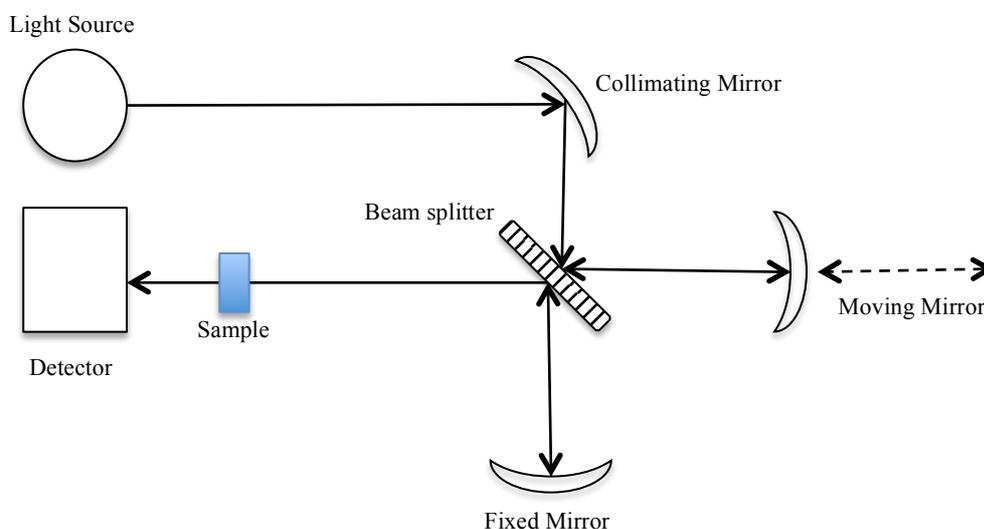


Figure 2-11 - Schematic representation of the key components of an FT-IR spectrometer and the Michelson interferometer.

Depending on the position of the moving mirror the combined beams of light may be in-phase on return to the beam splitter in which case constructive interference will

result creating an intense beam of light. If the distance travelled means the waves are out of phase, then destructive interference occurs resulting in a beam of light of weaker intensity. Measuring the intensity of light as the optical path length is varied, by moving the moving mirror, results in an interferogram.¹¹¹

Application of a Fourier transform to the interferogram results in the generation of a single beam spectrum, which is a plot of intensity against wavenumber. Single beam spectra are acquired initially with no sample present. This acts as a background spectrum and contains features associated with the instrumentation and the spectrometer atmosphere. Absorbance, or transmission, spectra of the sample are obtained by ratioing the single beam spectra of the sample and the background.¹¹¹

Fourier transform instruments offer a key advantage over other instrument designs, the high signal-to-noise ratio (SNR) of the observed peaks, with ratios of between 10 and 100 times greater than conventional spectrometer designs reported. Additional improvements in the signal-to-noise ratio can be achieved by increasing the number of scans averaged to give the final spectrum.¹¹¹

2.5 Chemometrics

Spectroscopic data collected in the near and mid infrared regions contains information about a variety of different components contained within the sample matrix. This contribution of various components to the resulting spectra often makes univariate analysis techniques inappropriate, so multivariate analysis methods must be employed.

2.5.1 Principal Component Analysis

The concept of principal component analysis (PCA) can be employed as a data reduction technique. In a simplistic two-dimensional arrangement both an X and Y co-ordinate can describe the points in a dataset. However using principal component

analysis the dimensionality of this dataset can be reduced to a single axis where each point is represented as a linear combination of the original variables (Figure 2-12).

The first principal component is drawn through the dataset so it describes the maximum variance observed in the samples and the points projected onto this line. Subsequent principal components describe increasingly less of the variance in the dataset and are positioned orthogonal to the previous principal components. The net result of this process is that both the substantial variance in the data and the variance that arises due to random noise are described by a number of principal components (usually equivalent to the number of original variables). However the principal components that contain variance due to random noise can be discarded, leaving a reduced dataset that only describes significant variation in the data.^{118, 119}

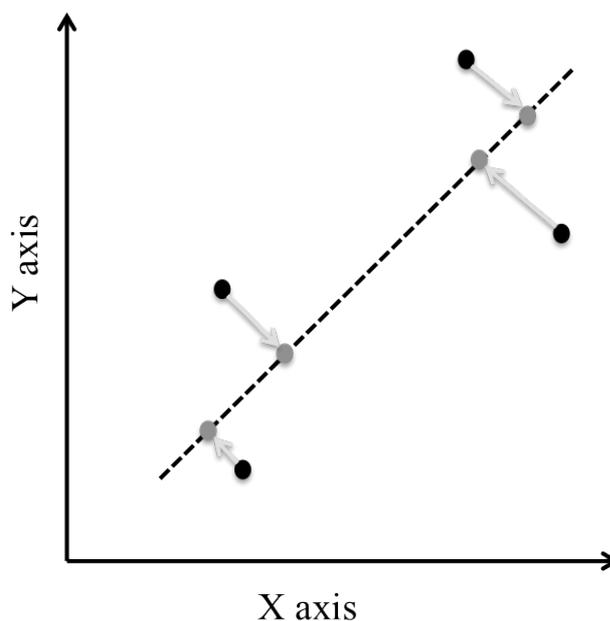


Figure 2-12 - Diagrammatic representation of data reduction by principal component analysis on a simple two-dimensional dataset.

Mathematically the original data matrix (in the case of spectra a matrix of absorbance measurements [n rows] made at various wavelengths [m columns]) is converted into two smaller matrices (T_k and V_k^T) as well as an error matrix (ϵ) according to Equation 9.

$$A = T_k V_k^T + \epsilon \quad \text{Equation 9}$$

In this case the matrix T_k represents the scores matrix, which gives information about the samples (or the spectra of the samples) and how they relate to each other. The other matrix V_k^T is the loadings matrix, which contains information about how the measured variables are related (the spectral regions exhibiting the variation). So each individual source of variance in the original data matrix (A) can be represented by the sum of the constituent principal components each of which has an associated scores and loadings matrix (Equation 10).^{119, 120}

2.5.2 Spectral Pre-processing

A variety of pre-processing techniques can be applied to spectral data prior to carrying out multivariate analysis or regression. These mathematical manipulations of the spectroscopic data were done with the intention of improving the quality of the spectroscopic data and ultimately enhance the constructed model to improve the accuracy of analyte concentration predictions.

2.5.2.1 Mean Centering

Assuming that the spectral data is represented by a numerical matrix where each column represents a measured wavelength, and each row the spectrum of a sample. The process of mean centering calculates the average of all the measured values for each column of the data matrix, and then subtracts this value from the measured value in each row of that column. Points in this data matrix therefore have no mean value associated with them and the point of the origin for the dataset has been shifted.^{119, 120}

$$A = T_1 V_1^T + T_2 V_2^T + T_k V_k^T + \epsilon \quad \text{Equation 10}$$

2.5.2.2 Smoothing & Derivatives

Improvements to the signal to noise ratio of the spectra can be obtained by applying a smoothing function to the spectral data. The process of spectral smoothing takes a point in the spectrum and examines a narrow window of the spectrum around this point. An average value, or polynomial expression, for this window is calculated and substituted as the value for that point, the window then moves along removing the oldest point from the window and adding a new point before repeating the process. Savitzky-Golay smoothing, which is a polynomial smoothing technique, is the most commonly employed smoothing technique.¹¹⁹

Reduction of baseline offsets and enhancement of spectral features, by narrowing and sharpening peaks, can be achieved by calculating a derivative of the spectral data. As with the process of smoothing the data a moving window is utilised to fit a polynomial expression to the data that falls within the window region. The coefficients of this polynomial expression are then utilised to calculate the derivative of the data for this point. Once a point has been determined the window is moved along in the same manner as with the smoothing operation and the process repeated.¹¹⁹

2.5.2.3 Other Pre-processing Methods

There are many other spectral pre-processing methods available to assist in enhancing the quality of the acquired spectra. Processes such as normalisation (setting the sum of squares for each spectrum to 1) and baseline correction (subtraction of an average of a region containing only noise from a spectrum) can be utilised. More complex mathematical techniques such as multiplicative scatter correction (MSC) and standard normal variate (SNV) transforms have also been reported as a mechanism for enhancing reflectance spectra.¹¹⁹

2.5.3 Partial Least Squares

Partial least squares (PLS) regression incorporates aspects of principal component analysis (PCA), which attempts to identify the maximum variation in the spectral data, with multiple linear regression (MLR), which attempts to link the spectral features with measured concentrations. As a blend of the two methods, PLS attempts to capture the variance in the data and correlate this with the measured concentrations.¹²⁰

Two different methods of the PLS algorithm can be utilised, PLS1 and PLS2. The PLS1 approach involves constructing a separate model for each of the quantified variables. With the PLS2 approach, a single model is constructed for all the quantified variables at the same time.¹¹⁹

When PLS regression is carried out both the data matrices (spectra and reference data) are reduced to their corresponding scores and loadings matrices by a process similar to PCA (Equation 11 and Equation 12).

$$X = TP^T + \epsilon \quad \text{Equation 11}$$

$$Y = UQ^T + F \quad \text{Equation 12}$$

The matrices P^T and Q^T represent the loadings or factors associated with the model, with F and E the residuals associated with the model, the PLS algorithm attempts to have these values as close to zero as possible.¹¹⁹ Although both the X and Y data matrices have been converted to their scores and loadings matrices this process was done independently, a relationship between the scores associated with the X block (T) and those obtained from the Y block (U) is developed by an iterative calculation.

$$U = TW \quad \text{Equation 13}$$

This results in a matrix of regression vectors (Equation 13) where U and T represent the scores matrices from the two PCA like stages and W the matrix containing the regression coefficients. This relationship can then be used for the estimation of unknown 'Y' values given the spectrum of the unknown.¹¹⁹

2.5.4 Artificial Neural Networks

Calibration approaches such as PLS and principal component regression (PCR) assumes the presence of a linear relationship between a measured parameter in a system and the intensity of a spectral response. This is generally the case for spectroscopic data according to the Beer-Lambert law.⁶³

Artificial neural networks (ANN) adopt a non-linear approach to the modelling process. Links between the measured parameter and spectral response are determined according to a series of equations. The issue with such neural networks is the “black box” nature by which the regression parameters are calculated.⁸³ Models based on ANN's are more complex than their linear counterparts and their efficacy can vary substantially depending on the number and quality of the spectra utilised in the training stages.^{63, 83}

2.5.5 Model Evaluation

The suitability of constructed models is usually assessed by comparison of the root mean square error of calibration (RMSEC) and root mean square error of prediction (RMSEP) values. Obtaining similar values for both RMSEC and RMSEP indicates a model that has a good predictive ability.⁶²

The RMSEC value gives an indication of the error between the spectroscopic data and the reference data for the calibration dataset. Values for RMSEC are calculated according to Equation 14, where 'N' represents the total number of samples/spectra in the calibration dataset and 'f' the number of factors included in the multivariate model.^{52, 62}

$$RMSEC = \sqrt{\frac{(y_{actual} - y_{predicted})^2}{N - f - 1}}$$

Equation 14

2.5.5.1 Internal Validation

Internal validation procedures can be utilised as a means of evaluating how well the constructed model is performing. The procedure can be useful for optimising the number of latent variables used in construction of the model but is no substitute for external validation approaches.

Leave one out cross validation (LOOCV) is a commonly employed method of internal validation. A model is constructed based on the calibration dataset but with the first sample excluded. This sample is then predicted using the constructed model. Another model is then constructed from the calibration dataset, this time however the second sample is excluded and used to test the model. The process is repeated until all samples in the calibration dataset have been excluded and tested against a constructed model.⁶² With this approach the model is essentially being evaluated on data that has been included within the calibration dataset and so doesn't challenge the model to the same extent as an external validation procedure.

2.5.5.2 External Validation

Models are generally challenged using an external validation dataset containing data that has not been included as part of the calibration dataset. The error between values predicted for the unseen dataset using the multivariate model and those obtained from the reference analysis method (RMSEP) is calculated using Equation 15. In this case 'N' refers to the number of unique samples/spectra contained within the validation dataset.^{52, 57, 60, 62}

$$RMSEP = \sqrt{\frac{(y_{measured} - y_{predicted})^2}{N - 1}}$$

Equation 15

2.5.6 Calibration Transfer

Predictive models based on spectroscopic data are generally instrument specific as a result of manufacturing parameters such as optical differences in the probes, and instrument specific spectral features such as artefacts.^{111, 121}

To ensure a model works effectively with different sets of instrumentation there are a number of options available. Constructing and maintaining a separate model for each instrument was a possible approach. This would result in the most accurate predictions however it is both time and labour intensive.¹²¹

As an alternative approach, a global calibration model could be constructed. This would incorporate samples acquired on all the spectrometers into the calibration model effectively modelling the variations in the spectrometers. One of the key disadvantages of this approach is the requirement for the model be updated if a new spectrometer is added, or if there is any maintenance carried out on the instruments that could result in an optical change.¹²¹

The third option for dealing with these instrument specific variations is to utilise the same calibration model on each instrument but apply a mathematical treatment to standardise the spectra to the one instrument on which the calibration model was built. Advantages of this approach over the other proposed methods include the time efficiency, ability to easily incorporate new spectrometers and the increased accuracy of predictions.^{121, 122}

Different approaches to the calibration transfer / standardisation process have been proposed; direct standardisation (DS), piecewise direct standardisation (PDS) and spectral space transformation (SST).^{123, 124}

3. Materials & Methods

Over the course of this investigation various analytical standards and reagent grade compounds were utilised. In all cases these materials were supplied by Sigma-Aldrich Ltd. (Dorset, UK) unless otherwise stated.

3.1 Standard Preparation

Analytical standards were utilised for the analysis and quantification of samples by the analytical reference methods employed, but also as a means of acquiring the pure component spectra of the various analytes of interest.

In many cases the compounds were commercially available and were purchased as 99% analytical grade purity. In some cases a commercial standard was not available and required a synthetic route to be followed to generate the compound that could be used as an analytical standard.

3.1.1 Commercially Available Standards

High purity standard materials were commercially available for the majority of analytes of interest. Stock solutions were prepared by weighing out the appropriate mass of the compound, dissolving in an aliquot of distilled water before transferring to a volumetric flask (Fischer Scientific, Loughborough, UK) and making up to the graduation mark.

Subsequent standard solutions for quantification of the analytes were prepared by pipetting an appropriate volume of the stock solution in volumetric glassware (Fischer Scientific, Loughborough, UK) and adding distilled water to the graduation mark to obtain a standard at the required concentration.

The concentrated stock solution was also utilised to obtain the pure component spectrum of each analyte of interest. Dilutions of these spectroscopic samples, if required, were prepared using volumetric glassware as previously described.

3.1.2 Non-commercially Available Standards

A number of compounds investigated were not commercially available, or were unavailable at the required purity. A suitable standard of tertiary-butyl glycine (TBG) was prepared from a re-crystallisation process. Other compounds of interest however required synthesis of the compound and subsequent purification by re-crystallisation.

3.1.2.1 Tri-methyl Pyruvic Acid

Commercially available tri-methyl pyruvic acid (TMP) material was of sufficient purity for use as a spectroscopic and analytical standard. The material was supplied by Ingenza Ltd. (Roslin, UK) in aliquots as required.

3.1.2.2 Tertiary-butyl Glycine

Standards of TBG were produced by re-crystallisation of the crude product obtained from the reductive amination and distillation procedure (3.3.2.1). An aliquot of the resulting solution was transferred to a round-bottomed flask and the solvent completely removed by rotary evaporation (Buchi, Oldham, UK). The minimum volume of distilled water was added under refluxing conditions to completely dissolve the crystals again. Once dissolved the solution was rapidly cooled in an ice bath causing the TBG to crystallise in solution due to the materials low solubility in water. These crystals were collected by filtration and an NMR spectrum obtained to check the purity of the material.

3.1.2.3 Compound B

A high purity commercial sample of this material was not commercially available so it was necessary to synthesise and purify the material based on an adaptation of the procedures described by Clark *et al.* (2001) (Figure 3-1).¹²⁵

Approximately 5 g the parent substituted pyrrole compound was placed in a 1 L round-bottomed flask and dissolved in 6 M hydrochloric acid. This solution was then refluxed for a period of eight to ten hours. The acid reflux process cleaves the bond joining the carbonyl and amine functional groups. After reflux the resulting solution was pre-concentrated using a rotary evaporator (Buchi, Oldham, UK).

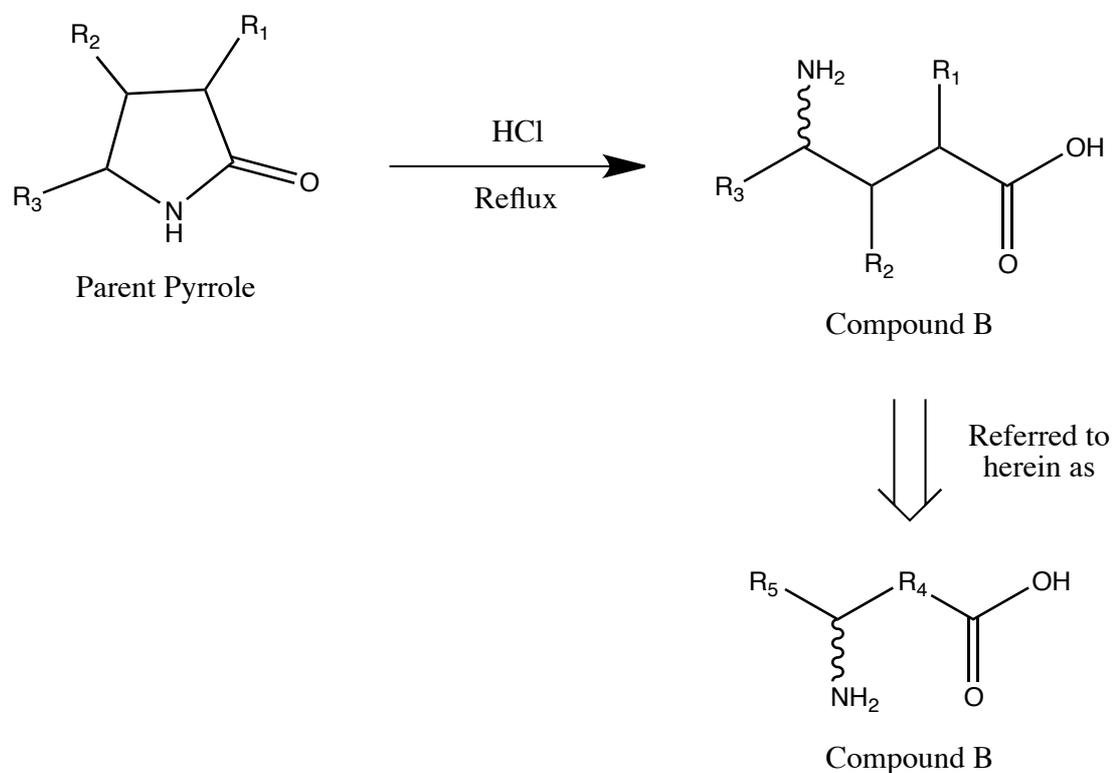


Figure 3-1 - Production of compound B for use as reference standard from the substituted pyrrole parent compound in an adaptation of the procedure described by Clark *et al.* (2001).¹²⁵

The desired product was extracted by flash chromatography using a strong cation exchange resin. A glass chromatography column was packed with Amberlite IR-120 plus strongly acidic cation exchange material (Sigma-Aldrich, Dorset, UK). This was washed initially with distilled water. An aliquot of the concentrated product from the reflux stage was then introduced onto the column and washed with at least one column volume of distilled water. The desired product remained trapped by the Amberlite column packing whilst the unwanted components passed through. A solution of 2 M ammonium hydroxide was passed through the column to elute 'compound B', with approximately 5 or 6 column volumes of eluent collected. The solvent was removed using a rotary evaporator to obtain compound B in crystalline form. Recrystallisation using the minimum volume of boiling methanol yielded the desired product, which was tested for purity by proton NMR.

3.2 High Performance Liquid Chromatography

Separation and quantification of the key analytes of interest in all cases was carried out utilising high performance liquid chromatography (HPLC) (Figure 3-2). The system employed was a modular system consisting of two Gilson model 306 pumps, a model 805 manometric module, model 811C dynamic mixer, model 234 auto-injector fitted with a 20 μ L volume sample loop, and model 832 temperature controller unit for the temperature controlled sample rack (Gilson, Middleton, USA). Column temperature was controlled using a Jones chromatography model 7971 column heater (Jones Chromatography, Hengoed, UK). This system was fitted with two detector units: a Gilson model 170 diode array detector (DAD) and also a Viscotek VE 3580 refractive index (RI) detector (Malvern Instruments, Malvern, UK). All the various components were controlled using the Unipoint software application [Version 5.11] (Gilson, Middleton, USA).

Individual chromatographic methods were developed for the separation and quantification of the various amino acids and keto acids investigated.

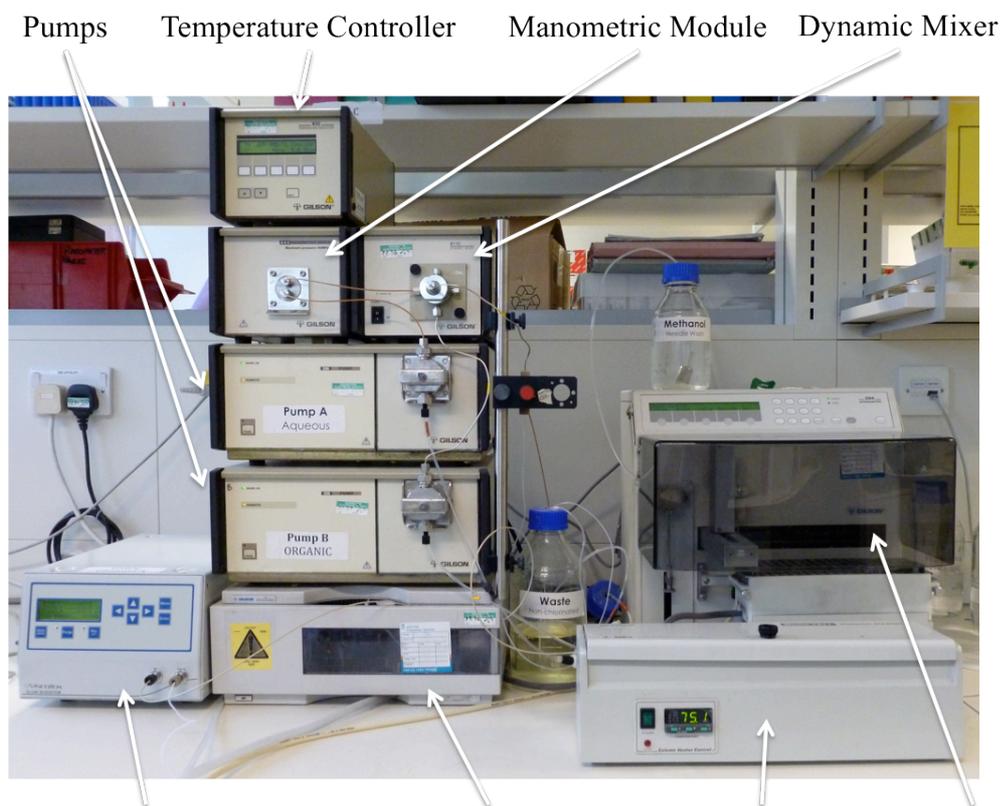


Figure 3-2 – Gilson modular HPLC system fitted with both refractive index and DAD detector systems.

3.2.1 Amino Acid Separation & Quantification

3.2.1.1 Derivatisation Procedure

The absence of a chromophore in the amino acid molecule required a derivatisation process to be carried out in order to detect the amino acids using the DAD detector system. The previously documented reaction between amino acids, *ortho*-phthalaldehyde (OPA) and 2-mercaptoethanol was selected as the derivatisation method (Figure 3-3).¹²⁶⁻¹²⁸ Based on the literature, this derivatisation process appeared to be compatible with a wide range of amino acids and would provide a consistent analytical procedure for most of the amino acids of interest.¹²⁸

To ensure consistency in the derivatisation process, but also due to time restrictions resulting from degradation of the derivatised complex¹²⁸, the process was carried out automatically by the auto-injector system prior to injection onto the chromatography column.

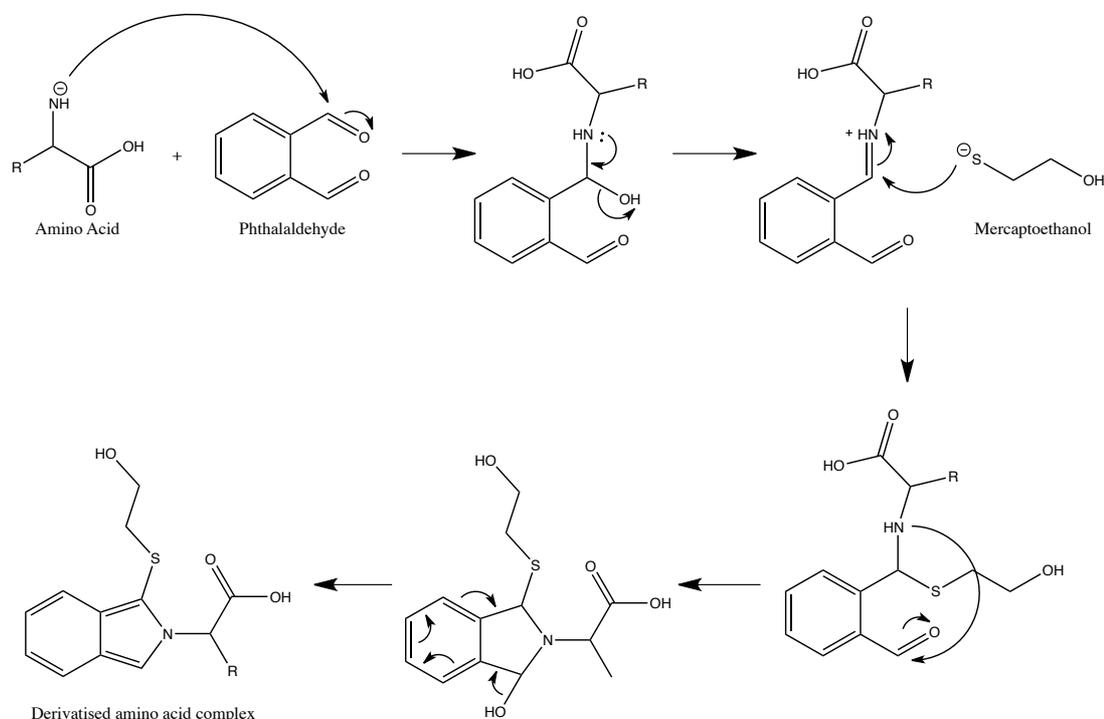


Figure 3-3 - Derivatisation of amino acid with *ortho*-phthalaldehyde (OPA) and 2-Mercaptoethanol to form chromophore containing compound allowing detection by DAD at $\lambda_{\text{max}}=230\text{nm}$ proceeding via mechanism suggested by Wong *et al.* (1985).¹²⁹

3.2.1.1.1 Preparation of Borate Buffer

A solution of 0.4 M borate buffer at pH 10 was prepared for use in the derivatisation process. Borate buffer was selected for the process as this buffer was reported as consistently resulting in a strong fluorescence¹²⁸ (so would be expected to have similar results with absorbance measurements), as well as being the buffer utilised by Ingenza for the method.

A mass of approximately 25 g of boric acid was weighed out and added to 500 mL of distilled water. The solution was stirred to dissolve as much of the solid boric acid as possible. A solution of 4 mol dm⁻³ potassium hydroxide was prepared by dissolving 224 g of potassium hydroxide in 1 L of distilled water. This solution was then added to the boric acid solution to adjust the pH to 10. As the pH increased the boric acid completely dissolved in the volume of distilled water. With the boric acid completely dissolved the liquid was transferred to a 1 L volumetric flask and the solution made up to the graduation mark with distilled water resulting in a 0.4 M borate buffer solution.

3.2.1.1.2 Preparation of Derivatisation Reagent

A derivatisation solution was prepared by weighing out approximately 20 mg of OPA using a 4-place balance (Mettler Toledo, Leicester, UK). This mass was then transferred to a 5 mL volumetric flask and 0.5 mL of HPLC grade methanol (Chromasolv, Sigma-Aldrich, Dorset, UK) added to completely dissolve the OPA. Using a micropipette (Gilson, Wisconsin, USA), a 20 µL volume of 2-mercaptoethanol was added and the solution made up to the graduation mark with the 0.4 M borate buffer solution. After mixing this derivatisation reagent was transferred into HPLC vials for use in the pre-column derivatisation procedure.

3.2.1.1.3 Derivatisation of Amino Acid

Derivatisation of the amino acid was achieved by mixing the amino acid sample, 0.4 M borate buffer at pH10 (this ensures the amine with a pKa in the region of 9.8 is ionised¹³⁰) and derivatisation reagent in a 1:2:2 ratio. To ensure consistency in the derivatisation, this procedure was carried out prior to injection of each sample. A 12 µL volume of the buffer solution was firstly aspirated followed by a 6 µL volume of sample and finally a 12 µL volume of derivatisation reagent. These solutions were mixed by aspirating and dispensing the 30 µL volume in the sample loop five times before being allowed to react in the sample loop for a four-minute period prior to

injection. Adopting this approach ensured each sample was consistently treated with the same amount of reagent and allowed to react for a consistent time period.

3.2.1.2 Chromatographic Separation

Separation of all the amino acids post derivatisation was carried out utilising a reverse phase C₁₈ chromatography column (Grace Davidson Discovery Sciences, Illinois, USA). Pre-column derivatisation was carried out according to the previously described procedure (3.2.1.1.3). In order to achieve the required separation and peak shapes of the various analytes of interest, different chromatographic conditions were utilised for the various amino acids under investigation.

3.2.1.2.1 Tertiary-butyl Glycine

The tertiary butyl glycine (TBG) amino acid was separated by a gradient elution method utilising methanol and distilled water containing 0.2% formic acid as the mobile phase. An initial composition of 65% aqueous, 35% methanol at a flow rate of 1 mL/min was held for two minutes. This was changed over a linear gradient between two and six minutes to 5% aqueous, 95% methanol, which was subsequently held for 6 minutes. The original composition was restored after 12 minutes and equilibrated until 14 minutes when the run was complete. Detection was using the DAD monitoring the complexes wavelength of maximum absorption at 340 nm, with a bandwidth of 10 nm, and referencing 500 nm, with a bandwidth of 40 nm.

3.2.1.2.2 Other Amino Acids

Separation and quantification of alanine, compound B and amino-butyric acid (ABA) was achieved using a gradient elution method with water containing 0.2% formic acid and methanol as the mobile phase. An initial composition of 80% water and formic acid, 20% methanol was held for a two-minute period. The methanol content was increased to 52% from two to four minutes over a linear gradient and held for a further minute and a half. From five and half to ten minutes the methanol content

was further increased to 90% and held at this level for eight minutes before being restored to the original conditions and allowed to equilibrate for subsequent injections.

3.2.2 Keto Acid Separation & Quantification

Quantification and separation of the keto-acid components was also carried out by HPLC. Some of these compounds did possess a chromophore and could therefore be detected using the DAD system, however the majority were detected using the refractive index detector system.

3.2.2.1 Tri-methyl Pyruvic Acid

The method employed for tri-methyl pyruvic acid (TMP) detection utilised a C₁₈ reverse phase chromatography column with detection via the DAD monitoring at 210 nm, with a bandwidth of 10nm, and referencing 400 nm, with a bandwidth of 40 nm. An isocratic method was employed with a mobile phase composition of 35% distilled water containing 0.2% trifluoroacetic acid (TFA), 65% methanol at a flow rate of 0.8 mL/min.

3.2.2.2 Other Keto-acids

All other keto acids of interest were analysed using a REZEX chromatography column (Phenomenex, Cheshire, UK). In this case the mobile phase was 0.005N sulphuric acid with a flow rate of 0.8 mL/min. Detection of the analytes with this method utilised the refractive index detector.

3.2.2.3 Organic Acids

Organic acids of interest, such a pyruvic acid and acetic acid, were also analysed using the REZEX chromatography column. For these analytes an isocratic mobile

phase composed of 0.01N sulphuric acid was utilised with a flow rate of 1 mL/min. Detection of these compounds was once again using the refractive index detector.

3.2.2.4 Method Validation & Error Estimation

Prior to use for quantification the developed HPLC methods were tested to ensure the responses obtained were linear, repeatable and also to estimate the error that was associated with any measurements made using these methods.

Linearity was assessed by preparing a series of calibration standard solutions of the analyte of interest at various concentrations. Each of these solutions was injected into the system in triplicate and the area under the peaks calculated. At least five calibration standards were prepared for each method to be validated. A linear line of best fit was plotted through the data points and the equation of the line was determined using the 'LINEST' command in Microsoft Excel (Microsoft, Washington, USA).

To assess the repeatability of the method an additional sample was prepared with a concentration that lay within the calibration range. This sample was injected for ten replicates and the area under the peak integrated and recorded. From these peak areas an assessment of the methods repeatability was determined.

The error associated with the fit of the calibration line through the data points was determined by calculating the root mean square error (RMSE) (Equation 16). Using the equation of the line of best fit through the data the concentration of each calibration sample was determined. This value was then subtracted from the true (known) value and squared for each sample. Summation of these values was divided by the number of injections and the square root taken to give the error associated with fitting the calibration line through the calibration data points.

$$\text{RMSE} = \sqrt{\frac{\Sigma(\text{Actual} - \text{Predicted})^2}{n}}$$

Equation 16

Errors arising due to variations between replicate injections also had to be taken into account when determining the total error associated with the method. A total of ten replicate injections of a sample not used for the calibration process, but within the calibration range were analysed using the method. Estimation of the error associated with these replicate measurements was calculated based on the confidence intervals. Peak areas from the ten replicate injections were used to estimate the concentrations using the previously determined regression equation. From these values the mean and standard deviation were calculated (Equation 17 and Equation 18), and the number of degrees of freedom (n-1) determined to obtain the T-Value at the desired confidence interval (Appendix I). The error value was then calculated using Equation 19 with the T-value obtained from an appropriate distribution at the 95% confidence interval.^{131, 132}

$$\bar{x} = \frac{\Sigma x}{n}$$

Equation 17

$$\sigma = \sqrt{\frac{1}{(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2}$$

Equation 18

$$\text{Mean} \pm \text{Confidence Interval} = \bar{x} \pm t \frac{\sigma}{\sqrt{n}}$$

Equation 19

The total error associated with the analytical method was quoted as the linear combination of the error due to lack of fit and the error associated with replicate injections calculated as per Equation 20.

$$\text{Combined Error} = \sqrt{((\text{Error 1})^2 + (\text{Error 2})^2)}$$

Equation 20

3.3 Biotransformation Processes

The biotransformation processes were carried out in bioreactor systems. These systems were chosen because they offered control, and in some cases data logging, of critical process parameters such as pH, dissolved oxygen levels, temperature, agitation and aeration rates.

3.3.1 Artificial Spiking

Due to the nature of the processes under investigation a stoichiometric linkage existed between the analytes of interest. This linkage presented issues of co-linearity in the resulting models of the system. To ensure that the resulting model is able to independently predict the concentration of all the analytes of interest it was necessary to break this co-linearity. Many approaches to breaking these relationships have been considered from generating a model using purely synthetic samples to retrospectively spiking samples with the analytes of interest.¹⁹ An alternative approach suggested was to utilise an experimental design to generate a series of mixtures in which there was no correlation between the analytes of interest.⁴⁶

In an adaptation of this approach an experimental design was utilised to determine the initial concentration of substrate and the introduction of a variable concentration product spike if appropriate. Breaking co-linearity in this manner also allowed for spiking of experiments where *in-situ* measurements were made.

Experimental designs were generated either as full factorial or partial factorial designs using the Design Expert software package [Version 8] (Stat-Ease Inc., Minneapolis, USA). The number of factors was determined by the number of key analytes of interest, with the levels chosen to encompass the concentrations expected during a typical replicate of the industrial biotransformation process.

3.3.2 Tertiary-butyl Glycine De-racemisation

3.3.2.1 Preparation of Tertiary-butyl Glycine

Racemic TBG for use in the biotransformation was produced from TMP via a reductive amination process (Figure 3-4). The high pressures of hydrogen gas required by the reaction meant the reductive amination was carried out in a Zipperclave model high-pressure laboratory batch reactor (Autoclave Engineers, Pennsylvania, USA). At this scale, the reductive amination reaction was carried out with a total volume of approximately 750 mL, which at the utilised loadings resulted in a solution with TBG concentration in the region of 1.6 mol dm^{-3}

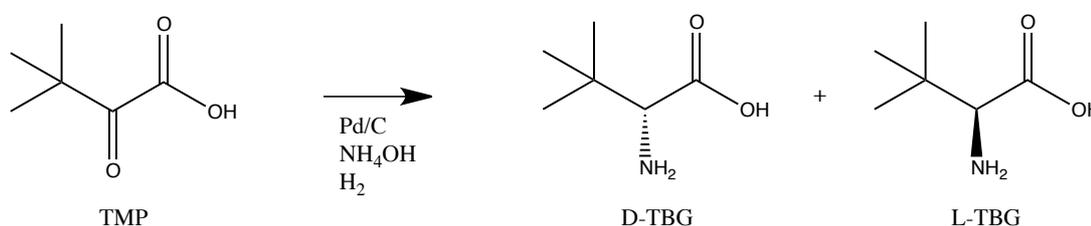


Figure 3-4 - Reductive amination process for the conversion of TMP to racemic TBG using ammonia solution under a pressurised hydrogen atmosphere catalysed by 5% palladium on carbon.

Using a balance approximately 165 g of solid TMP material was weighed out and transferred into the reactor vessel. To this 200 mL of neat ammonia solution at 35% (approximately 18 mol dm^{-3}) (Fischer Scientific, Loughborough, UK) was added, this would give a final ammonia concentration of 5 mol dm^{-3} . The reaction was catalysed using 5% palladium on carbon (Pd/C) (Degussa, Macclesfield, UK) of which approximately 8 g was weighed out and added to the reaction mixture. Distilled water was then added, approximately 550 mL, to take the total volume in the reactor up to the 750 mL volume. At this point the reactor vessel contained a solution with a TMP concentration in the region of 1.7 mol dm^{-3} , 5 mol dm^{-3} ammonia and 6 g/L palladium on carbon catalyst.

Having charged the vessel with reactant mixture, it was then positioned to form a seal and locked in place. At this point all the gas inlet lines were checked for integrity and to ensure they were closed. The water-cooling lines were also checked to ensure there were no leaks and that a cold water supply was available to the system.

With the vent line closed, the nitrogen supply line was opened and the reaction vessel pressurised with nitrogen gas. This gas was then released via the vent line and the process repeated a further twice to ensure the reactor vessel was purged of oxygen. Having completed three nitrogen purges, the hydrogen supply line was opened and the reactor vessel pressurised with hydrogen gas. Hydrogen gas pressure was maintained constant over the course of the reductive amination by the control system. Agitation in the system was set at a rate of 1500 revolutions per minute (rpm) and the temperature maintained at 70 °C.

The reaction conditions were maintained for a twenty-four hour period to allow a suitable yield of TBG to have formed. Supply of the hydrogen gas was stopped and the temperature and stirring rates decreased. Once the system had cooled sufficiently, the hydrogen gas pressure was released via the vent line. Three nitrogen purges were again carried out to remove any traces of hydrogen from the system prior to opening the reactor vessel.

Product from the reaction was filtered using a Buchner funnel and 0.2 µm filter paper (Fischer Scientific, Loughborough, UK) to remove particulates from the solution such as the spent Pd/C catalyst. This resulting solution was at a very high pH due to the presence of ammonia in the solution. Prior to use in the biotransformation stage, the pH of the resulting solution needed to be reduced to a pH in the range of 6 to 7.

A distillation process was utilised to reduce the pH of the TBG product. Using a quick-fit still head, condenser, 1L round bottomed flask and a heating mantle (Fischer Scientific, Loughborough, UK) a distillation setup was created. Product from the reductive amination was heated to boil off the ammonia solution. At various points during the process, distilled water was added to the solution to replace the

liquid that had been removed. Small aliquots of the solution were removed using a Pasteur pipette and the pH checked using pH paper. The distillation process was continued until the pH of the solution was approximately 7. At this point an aliquot of the TBG was removed and the concentration quantified by HPLC (3.2.1.2.1).

Based on the concentration determined by the HPLC analysis and the volume of TBG solution recovered from the distillation process, the appropriate volume of water was calculated and added to result in a TBG solution with a concentration of approximately 1 mol dm^{-3} . Again the actual concentration of the solution was determined by HPLC. At this point the TBG solution was suitable for use in the biotransformation process.

3.3.2.2 Biotransformation

Utilising a D-amino acid oxidase enzyme the de-racemisation of the amino acid TBG was carried out. The enzyme was contained within host *Escherichia coli* cells that had been provided in a freeze-dried state (Ingenza Ltd., Roslin, UK).

The process was carried out on a scale of approximately 550 mL in an Applikon bioreactor system comprising of an ADI-1030 bio console and ADI-1035 controller unit (Figure 3-5) (Applikon Biotechnology, Schiedam, Netherlands).



Figure 3-5 – Image of Applikon ADI bioreactor system illustrating the key components of the system.

As the cellular material was not growing during the biotransformation process, there was no requirement to maintain sterility within the system. However before starting the biotransformation process both the pH and dissolved oxygen probes required calibration. The pH probe was calibrated using buffer solutions at pH 4 and 7 (VWR, Leicestershire, UK). Initially the probe was placed in the low pH buffer solution and left for a period of time to equilibrate. Once a stable reading was obtained the calibration process on the instrument was started and the low value pH reading set. The electrode was then removed, washed with distilled water and placed in the buffer solution at pH 7. Again this was left until a stable reading was recorded and the reading stored by the instrument. When these readings had been saved the instrument then calculated a slope and offset that were used to calculate pH from the electrode readings.

Like pH, the dissolved oxygen probe also required calibration prior to use, however this was done when the initial substrate loadings had been measured out and placed in the reactor vessel.

Using the TBG solution of known concentration prepared from the distillation process (3.3.2.1) the initial substrate solution for the biotransformation was prepared. Approximately 330 mL of the racemic TBG solution and 210 mL of distilled water were mixed together in the reactor vessel. Volumes of TBG solution and water quoted were based on a final TBG solution concentration of 1 mol dm^{-3} , however the actual volumes used were calculated based on the concentration of TBG determined from the HPLC analysis. To this mixture 1 mL of polypropylene glycol (PPG) to act as an antifoaming agent and 2 mL of catalase from *micrococcus lysodeikticus* was added to the mixture.

This liquid solution was then placed in the reactor vessel and calibration of the dissolved oxygen probe carried out. The solution was heated to the temperature at which the biotransformation process would be carried out ($30 \text{ }^{\circ}\text{C}$). Oxygen free nitrogen (BOC, Manchester, UK) was introduced into the system by sparging at a flow rate of approximately 1vvm and the solution agitated at 300 rpm. Nitrogen was bubbled into the vessel until the dissolved oxygen probe returned a consistent signal that was then set as the zero value. The process was then repeated with air to obtain the high dissolved oxygen reading and this was set as 100%.

E. coli cells containing the DAAO enzyme made up the remainder of the reaction volume. A cell concentration of 50 g/L of the whole cells was required for the biotransformation process. Freeze dried cell material was however used for this process, so the mass of freeze dried enzyme was calculated as a mass per litre from the freeze drying factor and then this value adjusted for the required volume.

Samples were then removed from the reactor vessel at various stages during the biotransformation process. Aliquots of approximately 5 mL were taken via the sampling port, with the pH, temperature and dissolved oxygen values recorded for each sample. Once removed from the reactor these samples were heated for a three minute period at $100 \text{ }^{\circ}\text{C}$. The purpose of this step was to denature the enzyme, preventing any further conversion of the racemic TBG mixture between the point of sampling and analysis.

The biotransformation process was continued for at least a twenty-four hour time period, or until the dissolved oxygen levels returned to, or close to 100%.

3.3.3 Amino Butyric Acid De-racemisation

In common with the other de-racemisation processes under investigation, the de-racemisation of ABA was achieved by the selective oxidation of the D enantiomer to the corresponding keto acid leaving the L enantiomer untouched (Figure 3-6).

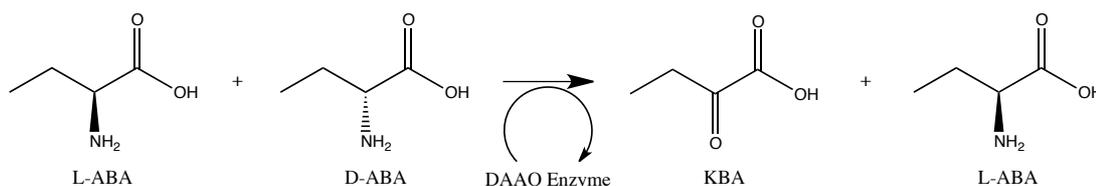


Figure 3-6 - De-racemisation of D/L-ABA using the freeze-dried DAAO enzyme.

Replicates of the process were carried out at a smaller scale in comparison with the de-racemisation of TBG process (3.3.2). In this case the total volume of the biotransformation reaction was 300 mL.

To accommodate this smaller volume a Braun Biostat Q (Sartorius Stedim Biotech, Aubagne Cedex, France) bioreactor system was utilised, which allowed up to four replicates of the biotransformation process to be carried out simultaneously (Figure 3-7).

In this case racemic ABA was commercially available, as was the KBA product of the biotransformation process. The substrate solution was prepared by weighing out the appropriate masses of ABA, and KBA if the system was being artificially spiked, into the reactor vessel and dissolving in 300 mL of distilled water.

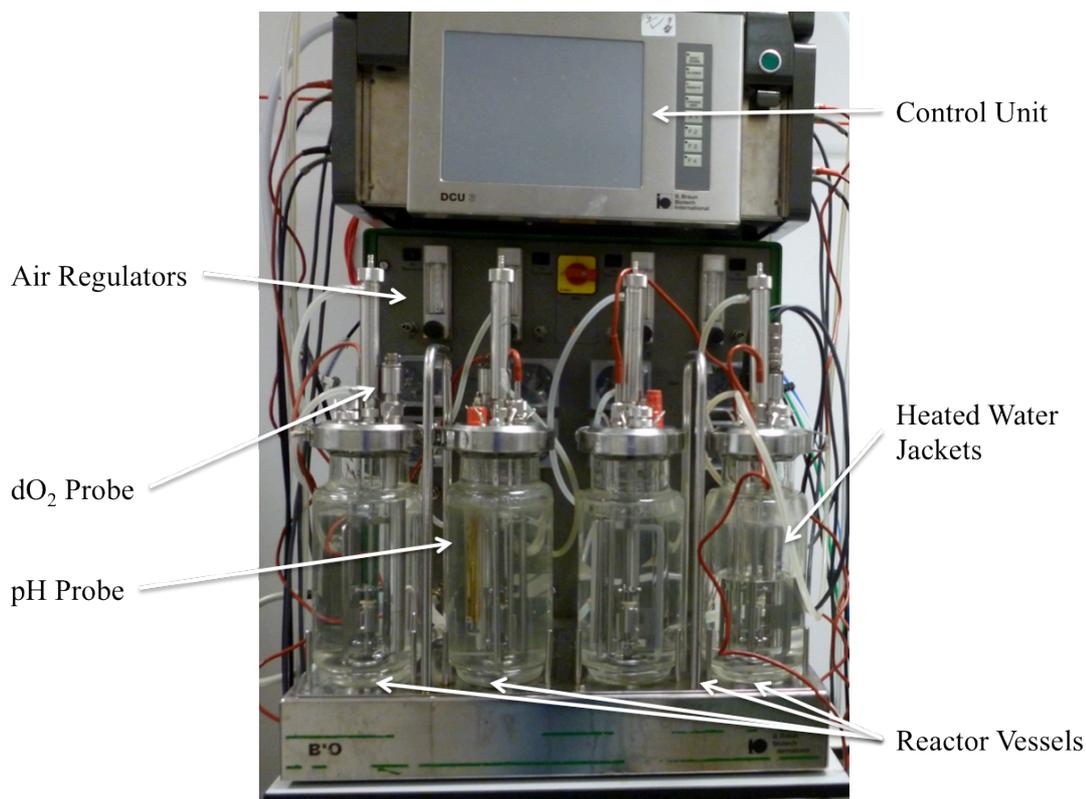


Figure 3-7 - Image of Braun Biostat Q bioreactor system illustrating the key components of the system.

Both the pH and dissolved oxygen probes for this system also needed to be calibrated prior to starting the biotransformation process. This was done following the same process with the buffer solutions and oxygen free nitrogen sparging as had been adopted previously (3.3.2.2).

Freeze dried cells of *Pichia pastoris* containing the DAAO enzyme were provided by Ingenza. A final wet cell density of approximately 7 g/L whole cells was required for the biotransformation, this approximately equated to 2 g/L of the freeze-dried cellular material. The appropriate mass of freeze-dried cells for the 300 mL volume was calculated and weighed out.

An initial sample of the reactor contents was taken as a reference point and the freeze-dried enzyme subsequently added. Sample aliquots of approximately 2 mL volume were removed at regular intervals over the course of the biotransformation

process. Once removed these samples were heated for a three minute period in an oven at 100 °C to prevent further conversion during storage.

The process was allowed to continue for a period of up to five days, with samples taken regularly over the first three days followed by a final sample on the fifth day prior to stopping the process. Samples were then stored at 4 °C until ready for analysis.

3.3.4 Alanine De-racemisation

The de-racemisation of alanine was utilised as the biotransformation process to investigate the use of *in-situ* mid infrared spectroscopy (3.4.2). Similar to the other systems investigated the D enantiomer of alanine in a racemic mixture was selectively converted to pyruvic acid leaving the L-alanine un-reacted (Figure 3-8).

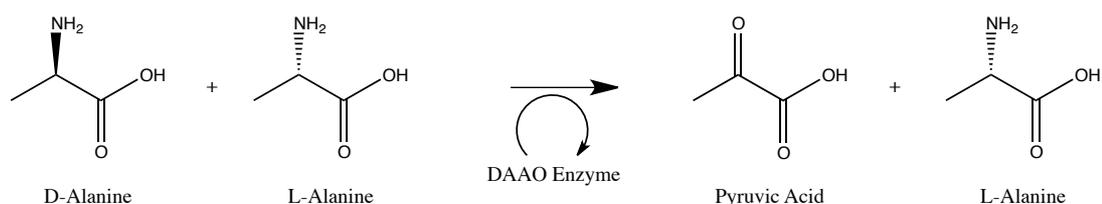


Figure 3-8 - De-racemisation of D/L-alanine to L-alanine and pyruvic acid using DAAO enzyme contained within a *Pichia pastoris* host cell.

In common with the other de-racemisation biotransformation processes investigated, the DAAO enzyme was contained within freeze dried *Pichia pastoris* cells. These were the same cells and enzyme utilised for the ABA biotransformation process (3.3.3) and again had a final cell density of 2 g/L (freeze dried cells).

A suitable amount of solid alanine, and pyruvic acid if the system was being artificially spiked, was weighed out to give a final concentration in the region of 1000 mMol dm⁻³ alanine. The material was placed in the bioreactor vessel and a suitable volume of distilled water added. For a 1L biotransformation process approximately 89g of racemic alanine was required. At this scale the reaction was

carried out in the Applikon bioreactor system comprising the ADI 1030 bio-console and ADI 1035 control unit.

Before the substrate, and any product spike, was introduced into the bioreactor the pH and dissolved oxygen probes were calibrated as had been previously described (3.3.2.2). An air background reference was also collected on the infrared spectrometer prior to the probe being fitted to the bioreactor.

Air was sparged through the system at a rate of 1vvm and the temperature set at 35 °C. When the system had stabilised the appropriate mass of freeze dried enzyme was added and the automated spectral acquisition process at ten-minute intervals started. Samples, of approximately 5 mL volume, were removed from the process at hourly intervals for the first eight to ten hours and then at regular intervals for the remainder of the process. As with the previous systems the samples were heated to denature the enzyme and then stored in refrigerated conditions until required for analysis.

3.3.5 In-situ Biotransformation

3.3.5.1 Preparation of Cell Bank

The *in-situ* biotransformation process utilised a strain of *E. coli* that had been genetically modified to express the transaminase enzyme. The strain utilised was BW25113 delta dad A ING 10183 provided by Ingenza Ltd. (Roslin, UK).

Solutions of LB Bullion (Merck, Darmstadt, Germany) containing 2% w/v bacteriological agar (Oxoid Ltd., Hampshire, UK) were prepared and autoclaved. Ampicillin, to give a final concentration of 100 µg/ mL, was dissolved in the minimum volume of water and filter sterilised before adding to the solution, which was mixed, poured into petri-dishes and allowed to set.

In a sterile environment, the master culture was plated out onto the LB agar plates and incubated overnight at 37 °C until the *E. coli* had grown sufficiently and a single colony could be extracted from the plate.

Inoculum media was prepared with the composition as described below (Table 1) in a 250 mL conical flask and the solution autoclaved prior to use.

Table 1 - Composition of salt solution, trace element solution and inoculum media used for the growth of *E-coli* strain.

Inoculum (200 mL)		Salt Solution		Trace Elements Solution	
Component	Volume (mL)	Component	Conc ⁿ . (g/L)	Component	Conc ⁿ . (g/L)
Salt Solution	40	(NH ₄) ₂ SO ₄	10	CaCl ₂ .2H ₂ O	0.5
50% Glucose	4	K ₂ HPO ₄	73	FeCl ₃	10.03
1M MgSO ₄	0.4	NaH ₂ PO ₄ .2H ₂ O	18	ZnSO ₄ .7H ₂ O	0.18
Trace Elements	0.4	(NH ₄) ₂ H-citrate	2.5	CuSO ₄ .5H ₂ O	0.16
Carbenicillin	0.2			MnSO ₄ .2H ₂ O	0.15
Sterile Water	155			CoCl ₂ .6H ₂ O	0.18
				Na ₂ EDTA.2H ₂ O	22.3

A single colony of *E. coli* cells was extracted from the LB plate and introduced into the sterile inoculum media. This flask was then incubated at 37 °C and agitated at 200 rpm until the optical density of the solution was approximately 1 AU at a wavelength of 600 nm.

Once the desired optical density had been reached 0.8 mL of the culture was added to a culture storage vial (Sigma-Aldrich, Dorset, UK) containing 0.2 mL of glycerol that had been sterilised. The process was repeated several times to create a cell bank, the solution was then frozen at -80 °C for use at a later time.

3.3.5.2 *E. coli* Fermentation

The fermentation process was carried out using a stainless steel BIOSTAT C-DCU bioreactor system (Sartorius Stedmin Biotech, Aubange Cedex, France) with a total maximum volume of 22 L and working volume of 10 L (Figure 3-9). This reactor was equipped with four internal baffles (1.5 cm x 57 cm) and three six-blade adjustable height Rushton turbine style impellers. Air was supplied via a circular annular sparger located at the bottom of the vessel with temperature control provided via a water jacket. Four ports located near the bottom of the reactor vessel housed the pH probe (Mettler Toledo Ltd., Leicester, UK), dissolved oxygen probe (Mettler Toledo Ltd., Leicester, UK), sampling port and near infrared spectroscopic probe (Foss NIRsystems, Maryland, USA).

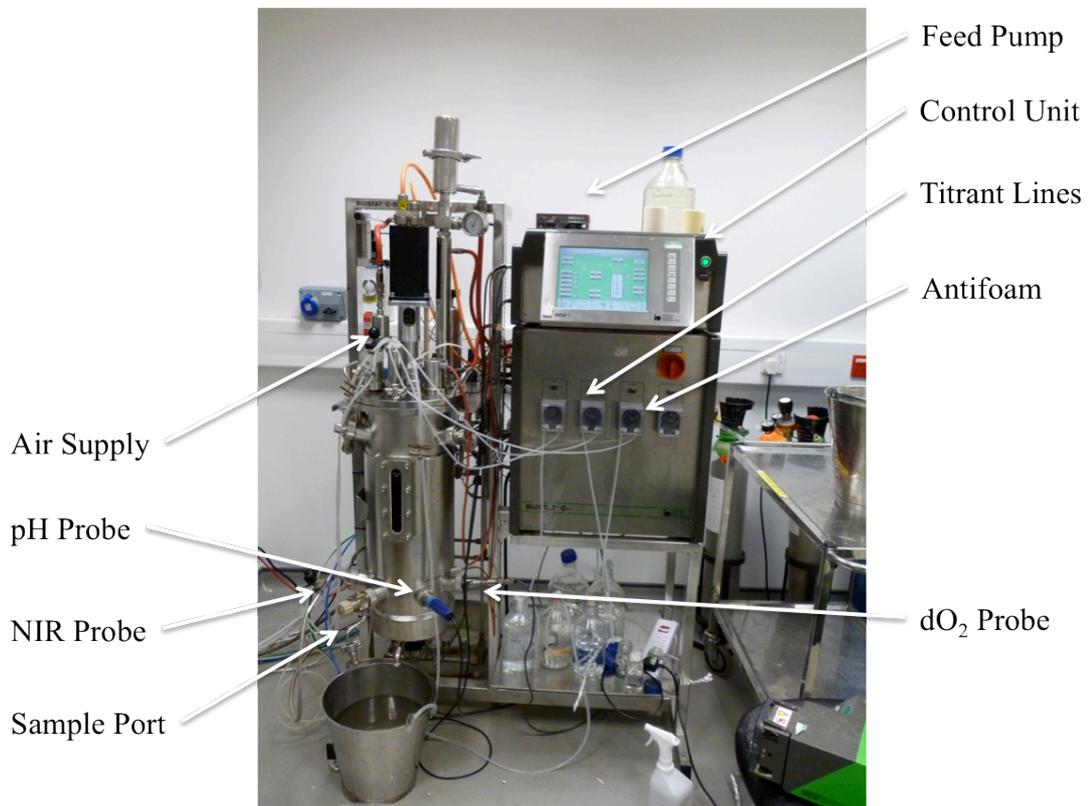


Figure 3-9 - Image of BIOSTAT C-DCU bioreactor system illustrating the locations of the key components and features of the bioreactor system.

Prior to use, the pH probe required calibration; this was achieved by measuring the probes response in a buffer solution at pH 7, followed by a buffer solution at pH 4 as described (3.3.2.2).

Growth media for the initial stage of the fermentation process was prepared, based on a total volume of 10 L with the components and concentrations detailed in Table 2. The media components (excluding those identified in Table 2) had the appropriate mass weighed out and were added to a volume of 9 L of distilled water. The solution was agitated until all the media components had completely dissolved. This was then transferred to the bioreactor system and sterilised by heating at 120 °C for twenty minutes. The glucose component was prepared by dissolving the mass of glucose in approximately 960 mL of distilled water, this was then autoclaved separately and added aseptically along with the media components that required filter sterilisation. Two conical flasks containing 300 mL of this initial growth media were also prepared and sterilised. These flasks were utilised to prepare the inoculum for the 10 L scale fermentation.

Table 2 - Batch media composition used for the growth of *E. coli*.

Component	Concentration (g/L)	Mass in 10L (g)
Ammonium Sulphate	2	20
Potassium Phosphate (dibasic)	14.6	146
Sodium Phosphate (monobasic)	3.6	36
Ammonium Citrate	0.5	5
Antifoam	0.1 mL	1 mL
Glucose*	11.9	119
Magnesium Sulphate (1M)*	2 mL	20 mL
Trace Elements**	2 mL	20 mL
Carbenicillin**	0.05	0.5

* Not autoclaved *in-situ*, sterilised separately and then added aseptically.

** Not autoclaved *in-situ*, filter sterilised and added aseptically

With the total volume of sterilised media in the reactor system the dissolved oxygen probe could then be calibrated. With agitation rate set at the highest rate expected during the course of the fermentation, and gas supplied at a flow rate typically expected during the fermentation oxygen free nitrogen was sparged into the media. When a stable reading from the probe was observed this was set as the zero calibration point. Compressed air was then re-introduced into the system until a stable reading was observed, which was set as the high point. Once calibrated, the agitation rate and airflow were set at 300 rpm and 10 Lpm respectively.

As this was a fed batch fermentation process, the glucose feeding solution, with glucose concentration of 550 g/L, was prepared at a volume of 2 L and sterilised in advance. To account for the increase in volume resulting from the addition of glucose a volume of approximately 1.1 L distilled water was measured out and the glucose (1.1 Kg) added in small aliquots of approximately 200 g. To completely dissolve the glucose it was necessary to heat the solution slightly using a stirrer hotplate.

Yeast extract to give a final concentration of 5 g/L when mixed with the glucose solution was weighed out and dissolved in 100 mL of distilled water. Both this solution and the glucose solution were autoclaved.

The glucose and yeast extract were mixed aseptically, with 50 mL per litre volume of 1M magnesium sulphate and 10 mL per litre volume of trace element solution also added aseptically and mixed well to give the feed solution.

Solutions of 1M sulphuric acid and 1M ammonia solution (Fisher Scientific, Loughborough, UK) were prepared by appropriate dilution of the relevant commercially available stock solution. These solutions were used as titrants to maintain the system pH at 7 during the fermentation.

Using the cell bank that had been prepared previously (3.3.5.1) two vials were removed and defrosted at room temperature. The contents of these vials were then

transferred into the two conical flasks containing 300 mL of the initial growth medium in an aseptic manner. These flasks were then incubated at 37 °C and shaken at 150 rpm until the culture returned an optical density in the region of 1 AU when measured at a wavelength of 600 nm. When the system had reached this point the inoculum was ready to be introduced into the bioreactor.

The bioreactor settings were verified prior to inoculation. In order to maintain the dissolved oxygen levels at 30% or above the agitation rate was cascaded between a minimum of 300 rpm and maximum of 900 rpm. Temperature was maintained at 30 °C, pH at 7 and aeration rate set at 10 Lpm (equating to an air supply of 1vvm).

Under an aerosol of ethanol the inoculation line was transferred into the 300 mL inoculum flask and the solution introduced into the bioreactor using a peristaltic pump (Watson Marlow, Cornwall, UK).

The fermentation was allowed to progress until the optical density of the culture fluid was in the region of 10 to 12 AU. At this point the cells became glucose limited and the feeding profile was begun.

When the feeding profile had completed the optical density of the culture media was in the region of 50 Au. Induction of the culture was then carried out by the addition of Rhamnose solution that resulted in a final concentration of 2 g/L in the reactor. Once induced the culture was left for a period of 17 hours before commencing the biotransformation stage.

3.3.5.2.1 Biomass Estimation

Estimation of biomass concentration in the bioreactor was determined by a measure of the dry cell weight. A 1 mL aliquot of the fermentation broth was transferred into a pre-weighed Eppendorf tube. This was then centrifuged at 10,000 rpm for a five minute period to form a pellet of the cellular material. The supernatant was removed and the process repeated using a 1 mL aliquot of distilled water. Following the

removal of the water supernatant the Eppendorf tube was dried in an oven at 105 °C for 24 hours and then allowed to cool in a desiccator. Once cool the tube was weighed again and the dry cell weight determined by subtracting the weight of the empty tube from this recorded mass. The process was carried out in triplicate with the quoted result being the mean value of the three measurements.¹³³

3.3.5.3 Transaminase Biotransformation

The biotransformation process for the production of compound B was carried out immediately after the fermentation process (3.3.5.2) and did not require the cellular material to be harvested or pre-treated.

In this process the enzyme swaps the amine functional group from the L-alanine substrate with the ketone functionality of compound A. The transaminase enzyme results in the formation of the compound B product in the S configuration (Figure 3-10).

An appropriate mass of solid L-alanine that would result in a final concentration of approximately 300 mMol dm⁻³ was weighed out. Compound A was utilised at a concentration in the region of 900 mMol dm⁻³, so again the appropriate mass of the substance was weighed out. Since the microorganism was only able to function within specific pH parameters it was necessary to pH correct the loading of compound A to 7 using 35% ammonium hydroxide solution.

When the 17 hour time frame after the point of induction had been reached, the cascading feature of the bioreactor was stopped, reverting the system back to an aeration rate of 10 Lpm (1vvm) and agitation rate of 300 rpm. All other parameters remained as they were during the fermentation process. Using a peristaltic pump the pH corrected compound A solution was added to the fermentation broth along with the L-alanine. From this point onwards a sterile environment was not maintained within the bioreactor system.

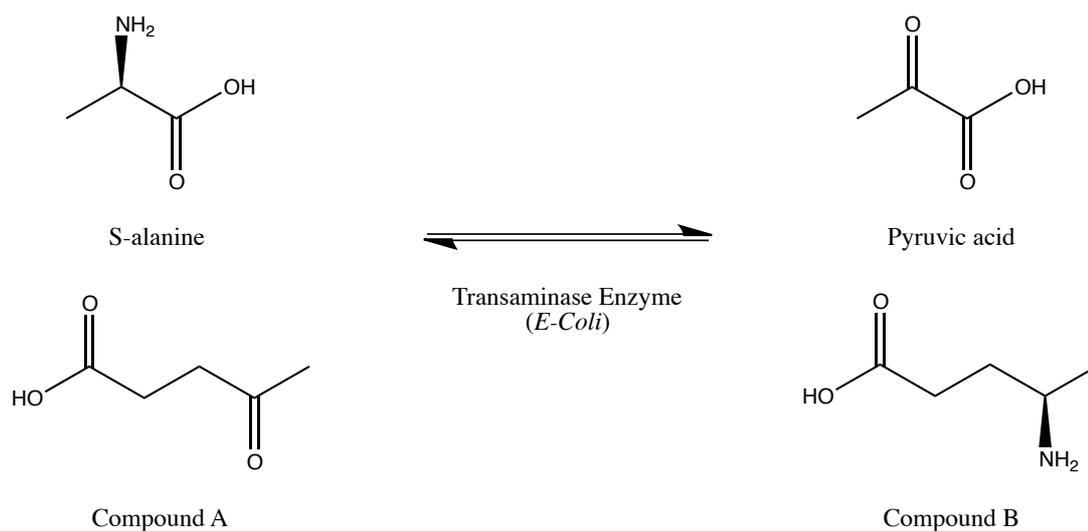


Figure 3-10 - Biotransformation process for the generation of compound B from L-alanine (absolute stereochemistry S-alanine) and compound A using a transaminase enzyme contained within the whole cell E-Coli resulting from the fermentation process (3.3.5.2).

Once the substrate had been added the biotransformation process was allowed to progress for a time period of approximately forty-eight hours. The time frame was monitored in conjunction with the dissolved oxygen levels, dissolved oxygen levels returning to 100% indicated the process had reached completion. Samples of approximately 20 mL in volume were taken at regular intervals over the course of the biotransformation. These samples were boiled at 100 °C to denature the enzyme and prevent any further generation of the product.

Near infrared spectroscopic measurements were made using the Foss XDS process analyser system fitted with a transfectance probe (3.5.2). Mid infrared spectroscopic measurements were made at-line using a Thermo avatar 360 system fitted with a zinc selenide ATR crystal (3.4.1)

When the biotransformation process had completed the material was destroyed using the *in-situ* sterilisation procedure.

3.4 Mid Infrared Measurements

3.4.1 At-line Mid Infrared

Acquisition of the at-line mid infrared spectra of the samples was carried out using a Thermo Nicolet Avatar 360 mid infrared spectrometer (Thermo Fischer Scientific, Massachusetts, USA). The instrument was controlled using the OMNIC software package (Version 5.2a). Sample presentation to the instrument was via the smartARK accessory (Thermo Fischer Scientific, Massachusetts, USA) fitted with a horizontal zinc selenide (ZnSe) attenuated total reflectance (ATR) crystal contained within a trough plate.

Using a Pasteur pipette, a thin layer of the sample material was placed on the crystal surface, ensuring that the entire surface of the crystal was covered and there were no air bubbles in the liquid film. The cover was then placed over the sample and clamped into place to prevent the evaporation of any volatile components that may have been present. The acquired spectra were reported as the average of thirty-two co-added scans of the sample and referenced against an air background. After acquisition each spectrum was saved in the Galactic (SPC) format for manipulation within another application at a later date.

3.4.2 In-situ Middle Infrared

In-situ spectroscopic measurements were made using an ABB MB3000 mid infrared FT-IR spectrometer system (ABB, Québec, Canada). The probe used for these measurements was a 12 mm diamond crystal ATR probe connected to the spectrometer by silver halide fibre optic cable (Fibre photonics, Livingston, UK). A probe launcher attachment was fitted to the spectrometer to allow coupling of the instrument and probe.

Spectroscopic measurements and instrument control were via the Horizon software application [Version 3.1.24.2] (ABB, Québec, Canada). Spectra resulted from the

average of 128 co-added scans of the sample material at a 4 cm^{-1} resolution and were referenced against an air background.

Spectra were acquired at ten-minute intervals over the course of the biotransformation process under investigation. Once acquired the spectra were saved in the Galactic (SPC) format and exported for manipulation and modelling.

3.5 Near Infrared Measurements

As with the mid infrared instrumentation, acquisition of the near infrared spectra could be achieved in either an at-line or *in-situ* manner.

3.5.1 At-line Near Infrared Measurements

At-line near infrared measurements of the sample materials were made using a Foss 6500 series near infrared spectrometer (FOSS NIRsystems, Maryland, USA). The instrument was equipped with a sample transport module (Foss NIRsystems, Maryland, USA) that allowed for sample presentation in fixed path length quartz cuvettes. Adopting this sampling approach allowed for rapid and simple optimisation of the sample path length for spectroscopic measurements made when the instrument was operated in both transmission and reflectance mode.

For most applications, a cuvette with a path length of 0.5 mm was sufficient to acquire suitable spectra without the dominating effects of water causing substantial interference. Prior to spectral acquisition, the spectrum of an air background was collected using an empty cuvette for use as a background reference.

In all cases the samples were placed in the cuvette using a Pasteur pipette and any air bubbles, which may cause further scattering of the incident light, removed by gently tapping the cuvette on the bench. The samples were placed in the guide and the spectra acquired. A total of thirty-two scans of the material were averaged and the resulting spectra referenced against the air background previously collected.

As before the spectra were saved and exported as a numerical spread sheet in comma separated variable (CSV) format that could then be imported into another application for manipulation and modelling at a later date.

Although the instrument and sampling accessory were able to acquire both the transmission spectra and the reflectance spectra, the same procedure was adopted regardless of the acquisition mechanism employed. Other than indicating the appropriate acquisition method on the instrument software, the acquisition mode could be switched between transmission and reflectance by either removing or replacing the blanking plate.

3.5.2 In-situ Process Measurements

Spectroscopic measurements were also made *in-situ* for a number of the processes under investigation. *In-situ* measurements were made using a Foss XDS process analyser (Foss NIRsystems, Maryland, USA); the system was also fitted with a multiplexing unit that allowed the one spectrometer to make measurements of a variety of systems. Spectra were collected using a transflectance probe (Foss NIRsystems, Maryland, USA) and connected to the spectrometer via a fibre-optic bundle. As with the at-line measurement system the probe design allowed for easy and rapid variation of the sample path length. In most cases, a gap of 0.5 mm resulting in an effective path length of 1 mm was sufficient to acquire reasonable spectra.

Measurements made *in-situ* were not referenced against an air background but against a NIST traceable reference material (serial number R99P0079). Due to the nature of the reference and the design of the probe the referencing procedure was carried out with a reflectance probe (Foss NIRsystems, Maryland, USA). A correction factor was then applied to compensate for the differences in the acquired spectra from the reflectance probe, used in the instrument calibration procedure, and the transflectance probe that was utilised for spectroscopic measurements.

The probe was mounted in one of the side sampling ports of the bioreactor. As these ports also housed the pH and dissolved oxygen probes, any sample measurements made should have been representative of the whole reactor contents and would not have had a significant impact on the mixing efficiency of the reactor.

Prior to mounting in the bioreactor the most appropriate path length for the sample under investigation had to be determined. This was done in an at-line fashion where the probe was placed in an aliquot of sample medium and the near infrared spectra acquired at various path lengths to determine the optimum. For most applications a gap of 0.5 mm, resulting in an effective path length of 1 mm, resulted in sufficient quality of spectra with all peaks of interest on scale.

Due to the probe design and size restrictions, *in-situ* measurements with the near infrared system were only carried out for the larger 10 L scale processes.

3.6 Spectral Manipulation & Modelling

Any infrared spectra acquired were exported from the various proprietary software applications and imported into the Matlab application [R2007b and R2009b] (The Mathworks, Massachusetts, USA) with the PLS toolbox add-in application [Version 4.0] (Eigenvector Research Inc., Washington, USA) for manipulation and modelling.

In most cases the acquired spectra were derivatised to enhance spectral features and reduce the effects of baseline drift.^{48, 134} Both the first and second derivatives of the spectra were calculated with Savitsky-Golay smoothing applied. In all cases the derivative was calculated from a second order polynomial and utilised a bandwidth of 21. All resulting models were constructed using the PLS toolbox add-in for Matlab.

4. Amino Acid De-racemisation

4.1 Process Overview

The initial biotransformation process under investigation was the selective de-racemisation of the amino acid tertiary butyl glycine (TBG) (Figure 4-1). Racemic amino acid was prepared by the reductive amination of tri-methyl pyruvic acid (TMP) using the procedure described (3.3.2.1)

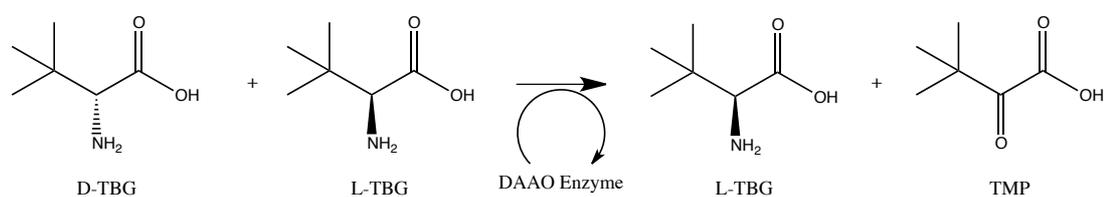


Figure 4-1 - Biotransformation process for the de-racemisation of TBG by selectively converting the D-TBG enantiomer to TMP leaving L-TBG untouched.

Currently there is very little real time or near real time process information obtained during the biotransformation process. Batch replicates of the process are typically allowed to progress for a given time period, with samples being taken at undefined intervals over the time course. Measurements made *in-situ* have been limited to traditional, well-established techniques such as pH, dissolved oxygen and temperature. Although dissolved oxygen levels can be utilised as an indicator of when the process is reaching completion (increased oxygen levels result from the decrease in enzyme activity reducing the oxygen demand to regenerate FAD (2.1)) no information about analyte concentrations can be obtained from these measurements. Information of this nature is typically obtained using off-line HPLC methods, introducing a time lag into the analysis between sampling and obtaining process information.

4.2 Aims & Objectives

This worked aimed to investigate the feasibility of utilising near and mid infrared spectroscopy as a method of monitoring this industrial biotransformation process in near real time.

The many reported successes of at-line near infrared spectroscopy with other similar optically challenging matrices (high concentrations of cellular material and multiple analytes with stoichiometrically linked concentrations) suggested that the near infrared region would be a suitable starting point to investigate for potentially monitoring biotransformation processes.^{16, 25, 27, 32, 35} These authors discuss applying near infrared spectroscopy to submerged culture bioprocesses, which not only exhibit chemical complexity such as stoichiometrically linked analyte concentrations and range of analyte concentrations but also constantly changing physical properties such as viscosity as a result of increasing cellular material. However despite success in overcoming these issues, the similar chemical structures of the species under investigation, differing only by a single functional group, meant the near infrared region may lack the required specificity and therefore that the mid infrared region should also be evaluated.

Initially an at-line approach was to be adopted due to the ease with which the samples could be presented to the spectrometer, and rapid nature by which instrumental parameters, such as path length, could be varied. Evaluation of both the near and mid infrared regions at-line will allow determination of which region, or if both, are suitable for use with this biotransformation process. Ideally moving to an *in-situ* sampling method will be the preferred option, as suggested by previous qualitative successes of infrared instrumentation with other biotransformation systems.⁷² However the at-line instrumentation available would be sufficient to allow the techniques abilities with this particular system to be evaluated.

Having determined the appropriate spectral regions, multiple replicates of the biotransformation process were to be carried out. Using the spectroscopic and

reference data from samples of these biotransformation replicates a quantitative model for the two key analytes of interest (TBG and TMP) was to be constructed. The resulting model was expected to undergo an external validation procedure using replicates of the biotransformation process that had not been utilised in the models construction. In addition the constructed model was also to be challenged using off-site replicates of the biotransformation process to further challenge the models abilities.

Due to the nature of the biotransformation system, with a known stoichiometric linkage existing between reactants and products, the contributors to the constructed model were to be evaluated to ensure that each analyte was being quantified independently and that co-linearity within the model was not an issue.

Despite this work being carried out at-line, the outcomes could be easily transferred to an *in-situ* system. Given the current approaches to process monitoring of these biotransformation processes, an at-line model represents a substantial improvement over the current off-line analytical methods (HPLC) employed.

4.2.1 Novelty

There has been limited application of infrared spectroscopy as a process-monitoring tool with industrial biotransformation systems. The reported applications of the technique with these systems has focused more on qualitative models.^{73, 95, 135} In some cases where quantitative models have been reported the samples have undergone a clean up stage, such as filtration, to remove much of the cellular material, which would present a challenging matrix to the spectrometer.^{95, 96} The presented work aims to construct a quantitative model for the biotransformation process using process samples that have not undergone any pre-treatment stages.

Many of the models previously published also do not interrogate the contributors to the model to ensure that co-linearity within the system has been adequately addressed. Trevisan *et al.* (2008) evaluate the loadings plots associated with their

models, however a separate PLS-1 model was constructed for each analyte of interest using the same spectral region it was unclear if co-linearity presented an issue for the quantification of each analyte.⁷² Models presented in this chapter were PLS-2 models, predicting both analytes simultaneously with examination of the regression coefficients giving a clear indication as to whether the issue of co-linearity had been addressed.

Additional novelty exists via the biotransformation process being carried out. Biotransformation processes involving de-racemisation have not been previously investigated with infrared spectroscopy. In other cases the reactant is consumed and product generated, however with the de-racemisation process at least 50% of the starting material will remain at the end of the reaction. Given the close structural similarities between analytes and the inability of the technique to distinguish between enantiomers these systems present a substantial challenge for monitoring.

4.3 Analytical Reference Method

Prior to carrying out any biotransformation replicates or collecting any spectra, it was necessary to ensure that a reliable and validated analytical reference method for quantification of both the analytes of interest (TBG and TMP) was in place. The method development process is detailed below.

4.3.1 Amino Acid

The method initially considered for separation and quantification of the TBG component was a direct transfer of the industrial method. Pre-column derivatisation of the amino acid was carried out using the OPA / mercaptoethanol derivatisation reagent (3.2.1.1). Conditions employed were based on a gradient elution method (program detailed in Table 3) at a flow rate of 1.5 mL/min using a C₁₈ reverse phase chromatography column. All mobile phase compositions are referred to as % volume/volume ratios.

Table 3 – Gradient elution method used by industrial partner for the separation and quantification of the total amino acid content in the samples.

Time (mins)	Buffer (% v/v)*	Acetonitrile (%)
0	95	5
4	75	4
10	55	45
10.5	95	5
12	95	5

* 10 mMol dm⁻³ potassium phosphate buffer solution.

Due to the high costs associated with the purchase of acetonitrile at the time of the experiments and the high volumes required, the method was modified and developed to utilise a more cost effective solvent.

4.3.1.1 Method Development

Acetonitrile was replaced by methanol as the organic solvent component of the mobile phase, and aqueous borate buffer replaced by distilled water containing 0.2% formic acid. The presence of formic acid ensured the amino acid remained protonated whilst passing through the column helping to ensure good chromatography and a symmetrical peak.

To provide a consistent method of pre-column derivatisation an automated script for the auto-sampler was written to carry out the derivatisation procedure. Initially the sample, buffer and derivatisation reagent were aspirated and dispensed into an empty vial. The solution was mixed repeatedly and left to react for a four-minute period before being introduced onto the column. Over the course of the method development process, the derivatisation procedure was refined to carry out this mixing step within the injection loop. Carrying out the derivatisation in this manner resulted in a more consistent derivatisation process and injection volume. When using vials the pressure associated with the action of dispensing liquid from the

needle caused droplets to form on the vial walls. Air bubbles in the sample solution from the mixing process and by attempting to aspirate this slightly reduced volume introduced inconsistencies into the injection process. By carrying out the derivatisation process and mixing within the sample loop these inconsistencies were reduced.

With a consistent derivatisation procedure in place, the mobile phase flow rate was reduced from 1.5 mL/min to 1 mL/min due to the high backpressure being observed on the system at the higher flow rate. Various mobile phase compositions were considered, from isocratic compositions through to gradient systems. The most effective separation was achieved with a high initial aqueous and moderate organic content that changed to a high organic content and very low aqueous content over the duration of the run. The best chromatography was observed with an initial composition of 65% aqueous, 35% organic held for two minutes before changing over a four minute period and linear gradient to 5% aqueous, 95% organic which was subsequently held for 6 minutes before reverting back to the initial conditions and equilibrating (3.2.1.2.1).

These conditions resulted in the best separation and best peak shape for the derivatised TBG peak. The resulting peaks were narrow, sharp and did not appear to exhibit any fronting or tailing, which was confirmed by calculation of the peak tailing factor (Equation 21) and peak asymmetry factor (Equation 22).¹²⁷

$$T_f = \frac{2B}{(A + B)_{5\%}} \quad \text{Equation 21}$$

$$A_s = \frac{B}{A_{10\%}} \quad \text{Equation 22}$$

Based on the measurements made (Figure 4-2), the peak returned a tailing factor of 1 and a peak asymmetry factor of 1. These values indicated that the resulting peak was symmetrical and did not exhibit tailing to any substantial degree.

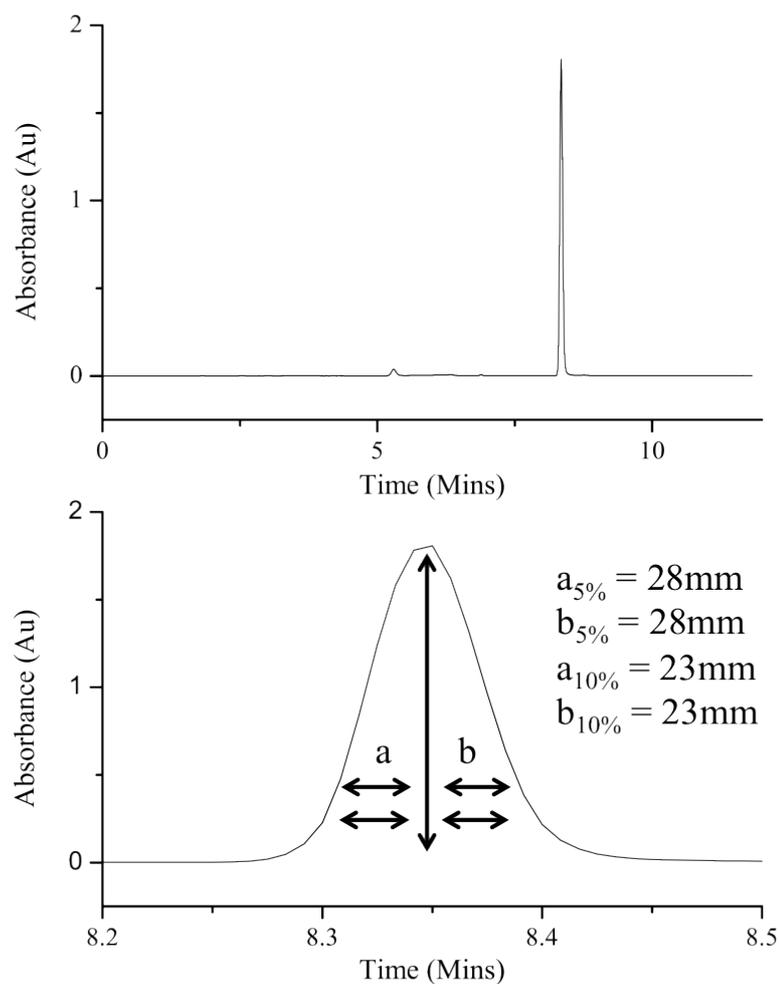


Figure 4-2 - Sample chromatogram of TBG analysed using the developed HPLC method and expansion of peak for the calculation of peak asymmetry factor and peak tailing factor.

4.3.1.2 Method Validation

To ensure that a linear response was being obtained for the amino acid over the concentration range of interest, a series of calibration standards were prepared and analysed using the developed method (3.2.2.4). A full set of calibration standards (covering at least 5 calibration points) were analysed and a new calibration equation calculated each time analysis of samples was carried out.

Samples were prepared between 2 mMol dm⁻³ and 10 mMol dm⁻³, at 2 mMol dm⁻³ increments in concentration, to give a five-point calibration curve. Each calibration sample was injected in triplicate and a plot of peak area response against concentration constructed (Figure 4-3), a linear line of best fit was drawn through the dataset and its equation determined using the 'LINEST' function in Microsoft Excel as previously described (3.2.2.4). Using this equation of the straight line, the concentrations of the calibration standards were predicted and compared with the actual values to determine the error associated with fitting the linear line through the data points (Table 4).

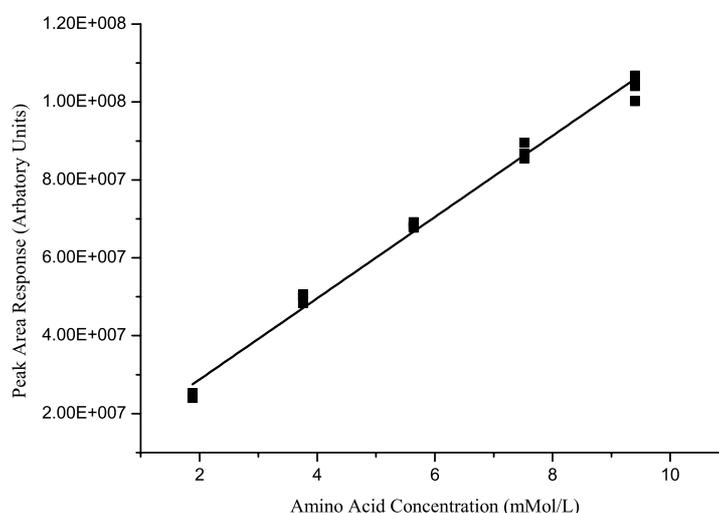


Figure 4-3 - Calibration curve with linear line of best fit through the data obtained for the calibration samples analysed using the developed TBG method. (Equation of straight line $Y=10429358.2x + 7892698$).

Table 4 - Calibration data used to determine equation of linear line of best fit through the data as being $Y = 10429358.2x + 7892698$, and calculation of predicted values to determine RMSE.

Concentration	Peak Area	Predicted	(Actual-Predicted)	(Actual-Predicted) ²
1.88	24167388	1.56	0.32	0.10
1.88	24017444	1.55	0.33	0.11
1.88	25199980	1.66	0.22	0.05
3.76	50610368	4.10	-0.34	0.11
3.76	50156760	4.05	-0.29	0.09
3.76	48428580	3.89	-0.13	0.02
5.64	67736640	5.74	-0.10	0.01
5.64	68188344	5.78	-0.14	0.02
5.64	69181448	5.88	-0.24	0.06
7.52	86752512	7.56	-0.04	0.00
7.52	89538696	7.83	-0.31	0.10
7.52	85582104	7.45	0.07	0.01
9.4	106814896	9.48	-0.08	0.01
9.4	104092624	9.22	0.18	0.03
9.4	100246392	8.86	0.54	0.30
Sum (Actual-Predicted)²				<u><u>1.00</u></u>

Since a five-point calibration curve was utilised with each sample injected in triplicate, the total number of samples was fifteen. Applying (Equation 16), the RMSE was determined as being $0.26 \text{ mMol dm}^{-3}$.

$$RMSE = \sqrt{\frac{\sum(\text{Actual} - \text{Predicted})^2}{n}}$$

$$\Rightarrow RMSE = \sqrt{\frac{1.00}{15}} = 0.26$$

In addition to the error associated with fitting the linear line through the data points there was an error associated with variations resulting from the replicate injections. Estimation of this component of the method error was particularly important given that a pre-column derivatisation stage was employed, and so variations resulting from the derivatisation process also needed to be taken into consideration. Ten

replicate injections of a standard solution were carried out and the concentration estimated using the equation of the linear line of best fit from the calibration curve (Table 5). The quoted value from these replicates was given as the mean value with the error quoted as the confidence interval (Equation 19).

A t-value of 2.26 was utilised, based on 9 degrees of freedom (n-1) at the 95% confidence interval (Appendix I). Using these values the mean concentration of the replicate injections and the associated error were calculated.

$$\mu = \bar{x} \pm t \frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow \mu = 2.06 \pm \left(2.26 \frac{0.09}{\sqrt{10}} \right)$$

$$\Rightarrow \mu = 2.06 \pm 0.06$$

Table 5 - Peak areas and calculated concentration of ten replicate injections to assess the repeatability of the developed method.

True Concentration (mMol dm⁻³)	Peak Area	Predicted Concentration (mMol dm⁻³)
1.88	28493786	1.98
1.88	30447978	2.16
1.88	30911048	2.21
1.88	29490266	2.07
1.88	29333060	2.06
1.88	29583148	2.08
1.88	30152694	2.13
1.88	28859192	2.01
1.88	29062752	2.03
1.88	27700142	1.90
Mean	29403406.60	2.06
Std Dev	949331.86	0.09

The quoted error ($\pm 0.27 \text{ mMol dm}^{-3}$) associated with the method was given as the linear combination of the error associated with the lack of fit from the calibration curve and the error associated with replicate injections. Relative to the median concentration standard used in the calibration range this represented an error of $\pm 4.8\%$ which was below the acceptable 5% error margin at the 95% confidence level.

$$\text{Total Error} = \sqrt{\text{Error 1}^2 + \text{Error 2}^2}$$

$$\text{Total Error} = \sqrt{0.26^2 + 0.06^2}$$

$$\text{Total Error} = \pm 0.27 \text{ mMol dm}^{-3}$$

4.3.2 Keto Acid

As with the amino acid, the method used for the separation and quantification of the keto acid and other impurities that may be present in the matrix utilised acetonitrile as the organic component of the mobile phase. The method was an isocratic method, with a flow rate of 1 mL/min and mobile phase composition of 60% water (containing 0.1% TFA) and 40% acetonitrile (containing 0.1% TFA). Separation was carried out using a C_{18} reverse phase column with the wavelength of detection set at 210nm.

In common with the amino acid, due to the high costs associated with the purchase of acetonitrile at the time, some development work was also carried out on the process transferred from the industrial methodology to find a more cost effective solvent.

4.3.2.1 Method Development

Switching to methanol as the organic component of the mobile phase, a variety of isocratic mobile phase compositions and flow rates were considered. Gradient elution methods were also investigated, however the best separation and peak shape were obtained with an isocratic mobile phase composition of 35% water with 0.2% TFA and 65% methanol, at a flow rate of 0.8 mL/min.

Using these conditions the resulting peak had a peak tailing factor of 1.05 (Equation 21) and an asymmetry factor of 1.12 (Equation 22). Both values suggested that the resultant peak was symmetrical and did not exhibit tailing or fronting.

4.3.2.2 Method Validation

Similar to the amino acid, a series of calibration standards and replicate injections were carried out to ensure a linear response was obtained, and to assess the error associated with the method.

A linear line of best fit was plotted through the data points and the equation of this line determined. As an estimation of the associated error, the RMSE for the calibration curve was determined as being $\pm 0.14 \text{ mMol dm}^{-3}$ and the error associated with replicate injections was found to be $\pm 0.03 \text{ mMol dm}^{-3}$ giving a combined error for the method of $\pm 0.14 \text{ mMol dm}^{-3}$. Relative to the median concentration of TMP used in the calibration data this represented an error of $\pm 1.9\%$ well below the acceptable upper limit at the 95% confidence interval. (Appendix II)

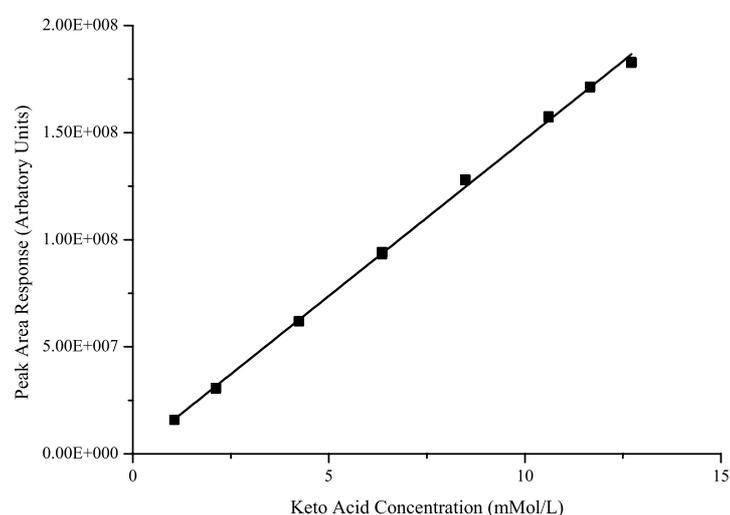


Figure 4-4 - Calibration curve for TMP, showing linear line of best fit through the calibration data.

4.4 Initial Feasibility Study

To determine how suitable infrared spectroscopic techniques were for monitoring the key analytes of interest (TBG & TMP) the pure component spectra of the two analytes were acquired. Solutions of both the TBG and TMP were prepared in distilled water at concentrations representative of those expected during the biotransformation process.

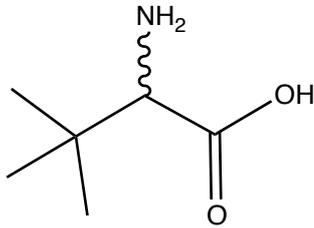
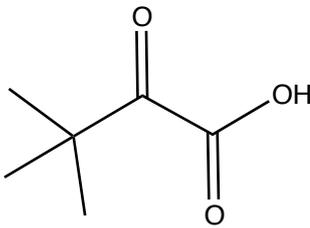
4.4.1 Near Infrared Spectroscopy

Initially the near infrared region was investigated. This region was selected based on the reported successes of the technique with other similar, and in some cases more complex, biological systems^{16, 23, 48} as well as being easily implemented for *in-situ* process measurements.

Since the two analytes of interest only differ by a single functional group, it was expected that there would be substantial similarities observed in the acquired spectra of the two components. Despite these similarities, there were some differences observed between the two molecules which, based on the theoretical regions where signals were expected (Table 6), should provide sufficient variation in the spectra to construct a quantitative model.

At-line measurements were made initially using the Foss 6500 system (3.5.1) using cuvettes with path lengths of 0.5mm, 1.0mm and 2.0mm. The best spectra (without any saturated signals) were obtained using the 0.5mm path length cuvette. The acquired pure component spectra of TBG and TMP were converted to the second derivative form (using a window size of 21) before being overlaid on top of the spectrum obtained for distilled water for comparison with the background matrix. Each of the regions (1st overtones, 2nd overtones and combination bands) were examined to identify where the spectral features of each analyte arose (Figure 4-5).

Table 6 – Theoretical locations of the signals expected to be observed in the pure component spectra of the TBG and TMP analytes.

Tertiary Butyl Glycine (TBG)		Tri Methyl Pyruvate (TMP)	
			
Functional Group	Regions (cm ⁻¹)	Functional Group	Regions (cm ⁻¹)
CH ₃	4167-4545 5882-6060	CH ₃	4167-4545 5882-6060
CH	4081-4385 5617-5952	COOH	5235-5319
NH ₂	4545-4761	CO	4920-5260 4360-4690
COOH	5235-5319		

Within the second overtones region (6250 cm⁻¹ to 9090 cm⁻¹) there was no significant variation observed between the spectra acquired of water and the pure component solutions of TBG and TMP. The lack of identifiable signals in this region wasn't unsurprising due to the expected weakness of any signals that may have arisen.

In the first overtones region some signals that arose from the TBG and TMP were identified. These signals were noted in the 5800 cm⁻¹ region and in the region just above 5900 cm⁻¹ and were common to both the analytes of interest. As these signals are present in both analytes, they would appear to correspond with the CH₃ functional groups common to both compounds, with only subtle variations observed in the peak apex. Of notable absence were the signals expected from the amine functional group of the amino acid (4545-4761 cm⁻¹) and also signals from the additional keto group present on the keto acid, most likely obscured by the dominating OH signals resulting from the water present in the matrix.

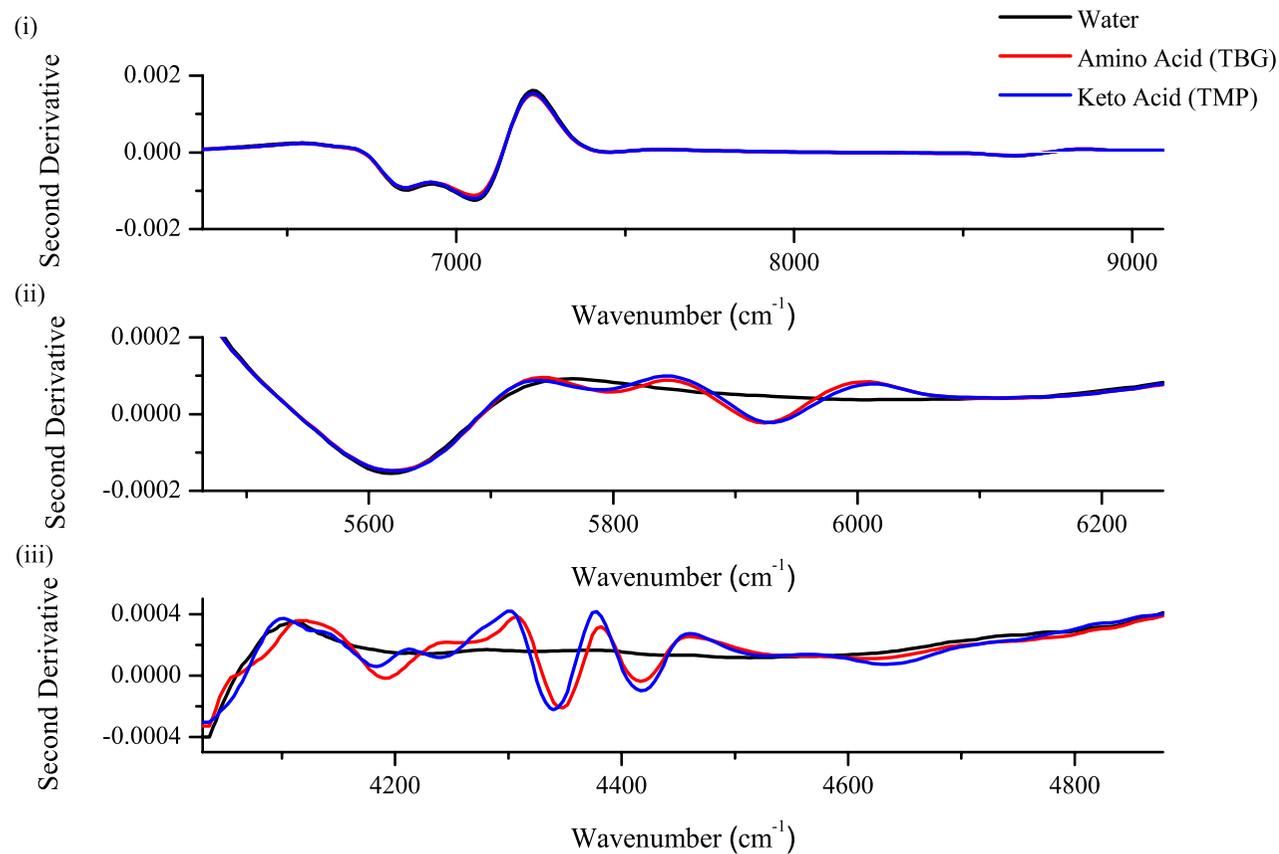


Figure 4-5 - Second derivative spectra of pure component solutions of the amino acid (TBG) and keto acid (TMP) plotted against the second derivative spectrum of distilled water. Regions were expanded to illustrate the (i) second overtones region (ii) first overtones region [dominant water signals have been removed for clarity] and (iii) the combination bands region.

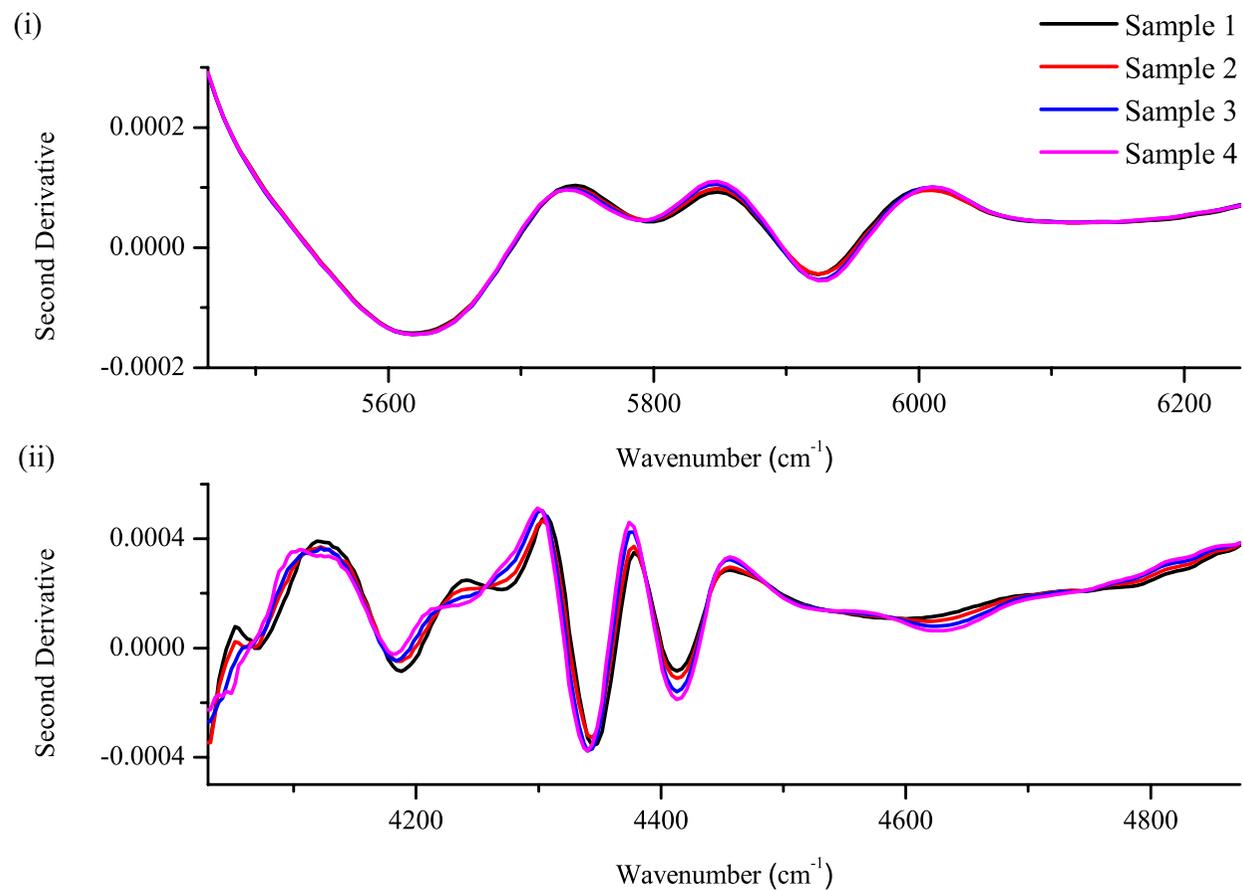


Figure 4-6 – First overtones (i) and combination band regions (ii) of near infrared spectra acquired of solutions containing a mixture of amino acid and keto acid at various concentrations (detailed in **Table 7**) typically expected during the biotransformation process.

Within the combination bands region a number of signals were observed, very few of which appeared to be unique to each analyte or displayed some degree of variation between signals arising from functional groups common to both analytes. The signals between 4200 cm^{-1} and 4500 cm^{-1} corresponded with vibrations of the CH_3 groups. Signals at and just below 4200 cm^{-1} can be attributed to the carbonyl functional group. This region provides points of difference between the two spectra, not only with a shift observed between the keto and amino acid, but also an additional peak at 4250 cm^{-1} present in the pure component spectrum of the keto acid.

These observations indicated that there was a very high degree of similarity noted between the pure component spectra of both the keto acid and amino acid. Whilst this was not unsurprising, a number of key spectral features specifically those arising from the amino acid functional group, appeared to be obscured by the dominant OH signals from the water background matrix. This brought into question the suitability of the near infrared spectroscopic region for monitoring this particular biotransformation process.

To determine if the near infrared region would be suitable, mixtures of the two compounds were prepared, with varying concentrations of each component, and the near infrared spectra acquired. TBG and TMP concentrations were within the concentration range theoretically expected during a typical process replicate and were determined according to an experimental design. To illustrate the spectral variations with concentration the infrared spectrum of four of these samples were plotted and superimposed (Figure 4-6). The approximate concentrations of TBG and TMP in each of these samples are detailed in Table 7.

Table 7 - Approximate concentrations of TBG and TMP of the samples whose infrared spectra were superimposed to illustrate the variations observed at concentrations typically expected during the biotransformation process.

Sample Number	TBG Concentration (mMol dm⁻³)	TMP Concentration (mMol dm⁻³)
1	650	0
2	500	125
3	400	250
4	250	350

Within the first overtones region the close similarities in the spectra made it difficult to identify spectral changes that could be attributed to one of the two key analytes when dealing with the mixture.

Within the combination bands, changes in the signals observed in the 4200 cm⁻¹ to 4300 cm⁻¹ region demonstrated changes in intensity that were in agreement with the known concentration profile of the samples (decrease in intensity of amino acid signals and appearance of keto acid related signals). Like the first overtones region, variations of the other spectral features identified with concentration were not as clear.

Based on these observations the suitability of the near infrared region for monitoring of this particular biotransformation process was questioned.

In other complex systems, such as fermentation broths, the weakness of the resulting signals has proved advantageous from a modelling perspective (2.4.2).^{17, 26} With the system under investigation however the weak signals, particularly in the first overtones region, meant changes in the analyte concentration did not translate to substantial changes in spectral intensity.

The main issue appeared to be in identifying spectral features that could be used to distinguish between the two key analytes of interest. Despite calculation of the second derivative spectra, to enhance and separate overlapping features as well as reduce baseline drift, it was still difficult to identify points of difference between the spectra of the two key analytes.¹⁸

The key point of difference between the spectra of the two analytes was noted in the combination bands region. As this is an industrial biotransformation process, the ideal scenario would be to apply the monitoring technique *in-situ*. Using the silica fibres commonly employed with near infrared systems results in the loss of the region below 4760 cm⁻¹. This is due to noise resulting from absorption of the light by the fibres.^{25, 47}

The close similarities in the spectra of the analytes of interest suggested it would be difficult to construct a quantitative model that was able to independently quantify both analytes for this particular biotransformation process. Any constructed model would rely heavily on the combination bands region, which would subsequently prevent the technique being applied *in-situ* due to limitations in the instrumentation, which prevents the transfer of light in this region via the fibre optic bundle.

Based on these results it would appear that the near infrared region is not suitable for monitoring this particular biotransformation process. This is, to an extent contrary, to expectations given the many successful reported applications of the technique to fermentation systems, many of which were inherently more complex in terms of content, light scattering properties and stoichiometric changes to the system.

Many of these reported successes with fermentation based systems however have linked the spectral features with either a physical property of the system, or to a number of key compounds within the fermentation broth that are structurally unique or present at high concentration.^{24, 32, 35} The mechanism by which signals in the near infrared region arise, mean the resulting signals are broad and weak. In fermentation systems this acts as an advantage due to the complex, highly concentrated and highly

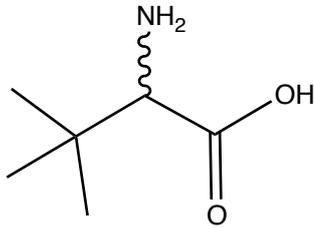
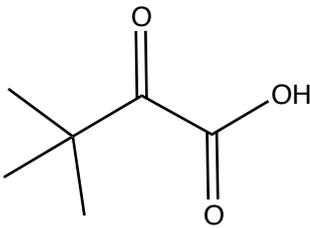
light scattering nature of the matrix.^{25, 26, 31, 81} However, in this biotransformation process a close similarity between the analytes of interest requires a higher selectivity than the near infrared region was able to provide. In this respect the mid infrared region may be better suited to monitoring this particular industrial biotransformation process.

Within the mid infrared region, the stronger, more defined signals should result in a more unique spectrum for each analyte, however the inclusion of the fingerprint region (below 1500 cm^{-1}), which by definition incorporates spectral features that are characteristic of a compound, would also mean sufficient variation in the spectra to construct a robust quantitative model.¹¹⁵

4.4.2 Mid Infrared Spectroscopy

The main signals expected to be observed in the diagnostic region for both the TBG and TMP were identified (Table 8). Due to the structural similarities between the two molecules the spectra were again expected to be very similar between both analytes. The most useful region of the spectra was likely to be the fingerprint region (below 1500 cm^{-1}) where the spectral pattern observed was expected to be unique for each analyte.

Table 8 - Theoretical regions where signals from the TBG and TMP were expected to be observed.

Tertiary Butyl Glycine (TBG)		Tri Methyl Pyruvate (TMP)	
			
Functional Group	Regions (cm ⁻¹)	Functional Group	Regions (cm ⁻¹)
CH ₃	1370-1390 2850-2960	CH ₃	1370-1390 2850-2960
CH	2880-2890	COOH	2500-3000 1700-1725
NH ₂	1560-1650	CO	1705-1725
COOH	2500-3000 1700-1725		

The mid infrared spectra of pure component samples of amino acid and keto acid at concentrations typical of the biotransformation process were acquired. Samples were measured at-line using the Thermo avatar spectrometer and the smartARK accessory fitted with a 45° zinc selenide ATR crystal trough plate (3.4.1).

When considering the full spectral region the aqueous background matrix presented challenges for the system. The Hydroxyl group signals, which were present in both the sample and background matrices, dominated the spectrum. These were then further broadened due to hydrogen bonding.¹¹⁵ Within this diagnostic region very few signals that could be attributed to a specific functional group within the analyte molecules could be identified due to these dominant signals.

Despite the apparent lack of signals in the diagnostic region, the fingerprint region appeared to exhibit some spectral features that could potentially be useful for constructing a calibration model for the system.

To try and enhance the spectral features obtained and to reduce the noise and baseline drift the spectra were derivatised, in this case the second derivative was utilised again with a window size of 21. A plot of the second derivative pure component spectra was then examined, focusing on the fingerprint region as opposed to the diagnostic region (Figure 4-7).

Close similarities in the structure of the analyte molecules again meant that there was a significant amount of overlap in the spectral features of both analytes. Despite this there were some regions where a shift in the peak position was observed or indeed unique features attributable to each analyte that could be identified.

Overlapping features between TBG and TMP were noted in the approximate region of 1050 cm^{-1} and in the area $1350\text{-}1450\text{ cm}^{-1}$ (Figure 4-7). Although these peaks overlap, the overlap observed was less than the comparable signals in the near infrared region, as a result of the narrower, more intense signals that arose.^{52, 54}

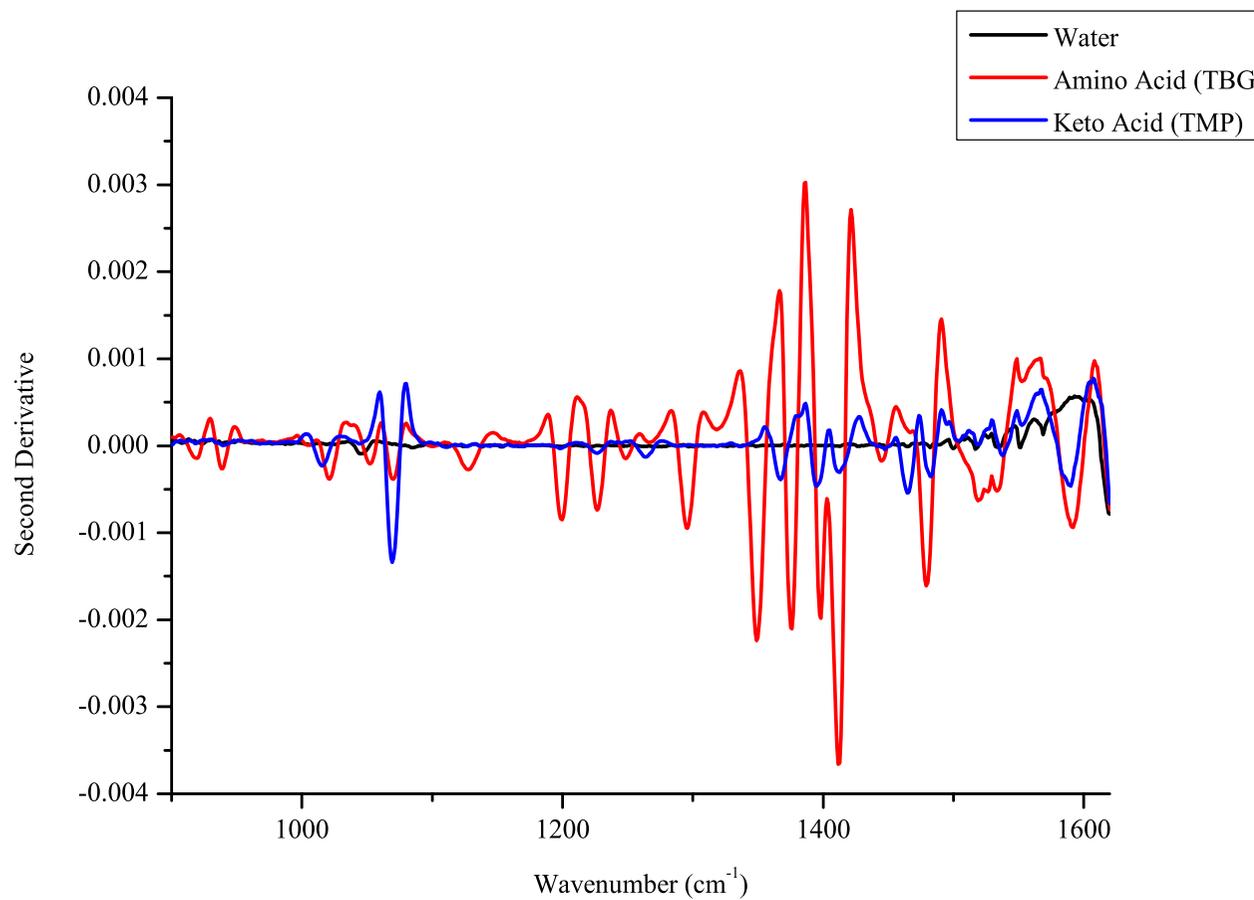


Figure 4-7 – Fingerprint region second derivative pure component spectra of the two key analytes of interest and water background matrix compared to identify useful regions for use in construction of a multivariate model for the biotransformation process.

A number of important regions that resulted in spectral features unique to each of the two analytes of interest were also observed. Importantly the amine functional group gave rise to a clear signal in the 1500-1550 cm^{-1} region, which is one of the key functional group differences between the two analytes and had been obscured by the water peaks in the near infrared region.

In the 1250-1280 cm^{-1} region unique signals for both the TBG and TMP compounds were identified. The keto acid gave rise to a signal at the higher end of this region that was separated from the signal of the amino acid observed more towards the 1250 cm^{-1} region.

Other points of difference in the spectra were noted between 1000-1050 cm^{-1} . These signals exhibit a substantial degree of overlap, however there was clearly a shift in the signal position between the two molecules of interest.

Similar to the near infrared region, these identified spectral regions and their behaviour were examined when the analytes of interest were presented as a mixture at varying concentrations typical of the biotransformation process. The approximate concentrations of TBG and TMP in each sample were the same as those prepared when investigating the near infrared region and are detailed in Table 7.

When considered as a mixture of the two analytes it was still possible to identify the spectral features of note in the identified regions. As the concentration of the samples was varied according to Table 7, the regions identified as arising from the TBG decreased in intensity (notably the signals in the regions 1500-1550 cm^{-1} and 1150 cm^{-1}) with those from the TMP increasing in intensity (**Figure 4-8**). Of particular note was the changes observed in the signals in the 1250-1300 cm^{-1} region, where there is a clear decrease in the TBG signal and increase in the TMP signal as the concentration of each analyte was varied in a manner representative of that expected to occur during the biotransformation process.

Based on these observations it would appear that the mid infrared region was more suited to monitoring this particular industrial biotransformation process. Distinct and overlapping spectral features attributable to each analyte of interest were identified, and variations in these features with concentration were observed.

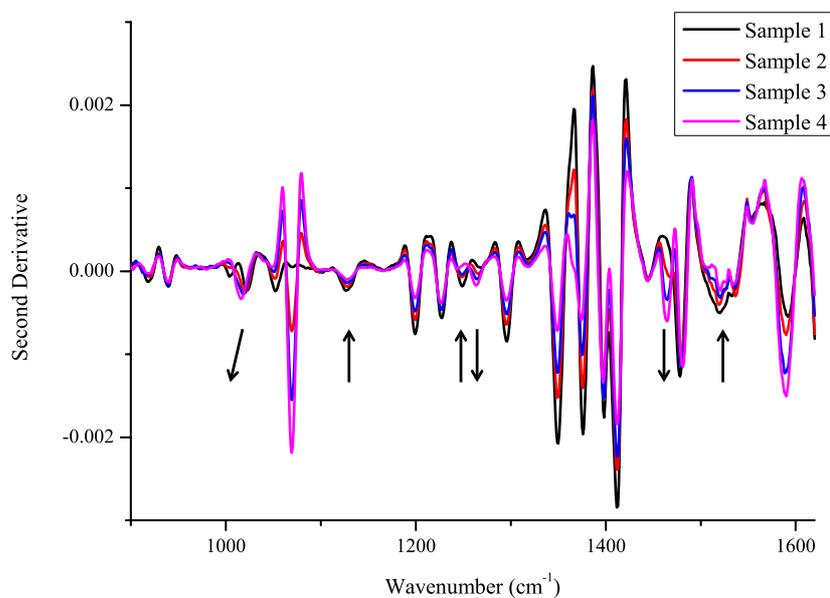


Figure 4-8 – Fingerprint mid infrared region of samples containing a mixture of TBG and TMP at varying concentrations in a manner representative of the biotransformation process as it progressed to identify regions that exhibited spectral changes for each analyte with concentration. Arrows denote spectral features that demonstrate an increase or decrease in signal intensity with concentration as detailed in Table 7.

4.4.3 Experimental Design

Prior to carrying out any scale replicates of the biotransformation process, a series of synthetic sample mixtures of the two key analytes of interest were prepared according to a simple 2 factor, 4 level (2^4) full factorial experimental design.

This approach aimed to refine and confirm that the identified regions could be utilised to construct a model for the system.

Raw spectra were converted to the second derivative spectrum and the appropriate region ($900\text{-}1620\text{ cm}^{-1}$) selected and extracted. Quantification of TBG and TMP in the synthetic samples was carried out using the described HPLC methods (3.2). Using the spectral data and reference data a partial least squares (PLS) model was constructed. This was a single model constructed for both the analytes of interest (PLS 2) based on the second derivative spectra. Both the spectral data and reference data were mean centered, with no additional pre-processing techniques applied.

Leave one out cross validation (LOOCV) was applied as an internal validation procedure to give a rough indication as to how good or bad the model was predicting the analyte concentrations. This approach is not ideal since the model was being tested using spectra that had essentially been used in its construction.⁴⁹ However, since at this stage the investigation was more concerned with the initial feasibility of the process the internal validation procedure employed was sufficient.

Two latent variables were selected as the optimum number for the model, determined from the plots of RMSEC and RMSECV for both TMP and TBG plotted against the number of latent variables. This model returned a RMSEC of 28.5 mMol dm^{-3} for the TBG and 4.1 mMol dm^{-3} for TMP, with RMSECV values of 30.8 mMol dm^{-3} and 4.4 mMol dm^{-3} respectively.

Examination of the plots of measured value (from the reference data) against predicted value (from the leave one out cross validation) indicated that the points appeared randomly scattered around the central diagonal, which suggested bias within the model was not an issue (Figure 4-9).

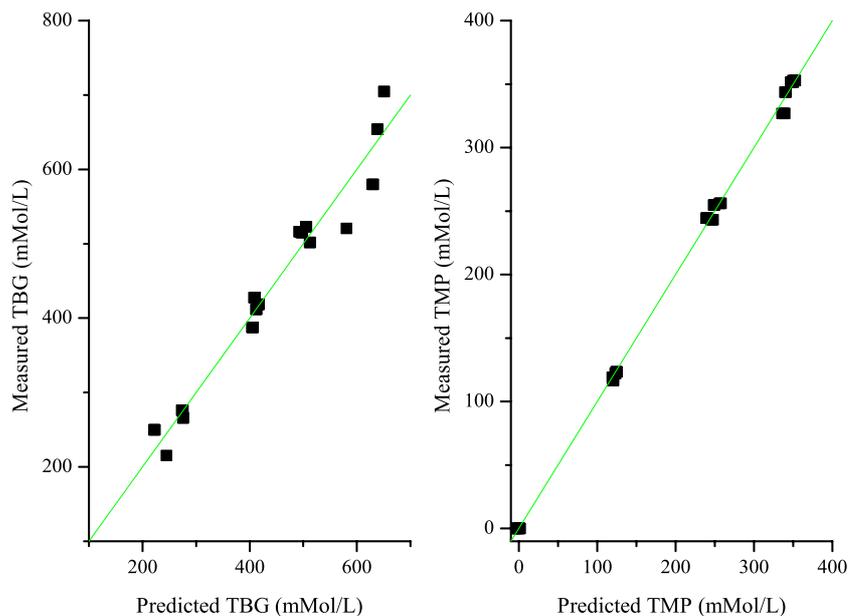


Figure 4-9 - Plots of measured concentration (mMol dm^{-3}) (from reference analysis) against predicted concentration (mMol dm^{-3}) (from leave one out cross validation) for both the analytes of interest in the synthetic experimental design model constructed using the selected fingerprint region of the mid infrared spectra.

Predictions for both the TBG and TMP were in good agreement with the measured values from the reference analysis. In this case however the samples were synthetic, and so were a relatively clean sample matrix with any stoichiometric link between the analytes broken from the outset. The model did however illustrate the feasibility of using mid infrared spectroscopy to monitor these key analytes in an industrial biotransformation process.

4.5 Calibration Model

Three batches of TBG were prepared from TMP by the reductive amination procedure described (3.3.2.1). Since the material was produced in a batch process, and that conversion of the TMP substrate would never reach 100% efficiency, the concentration of racemic TBG (and residual TMP) varied from batch to batch. Using the estimated concentration from the HPLC analysis, the appropriate volume of

substrate to result in a final racemic TBG concentration of 600 mMol dm^{-3} was measured out and added to the bioreactor. All other components were added as described (3.3.2.2) and made up to the final volume with distilled water.

Replicates of the biotransformation process were subsequently carried out. A total of five process replicates, utilising three batch preparations of racemic TBG, were carried out. Samples were taken at regular points during the biotransformation process and the mid infrared spectra of each sample acquired (3.4.1). Quantification of the total TBG and TMP content in the samples was also determined using the developed and validated HPLC methods (3.2.1.2.1 & 3.2.2.1).

During the biotransformation process the dissolved oxygen levels in the system were monitored as an indicator of how the biotransformation was progressing. Following the process in this manner was possible because the DAAO enzyme requires energy input from FAD generating $\text{FAD}^{\cdot-}$, which reacts with the oxygen to regenerate FAD (2.1) High levels of dissolved oxygen in the medium indicated that FAD is no longer being utilised and that the biotransformation process had reached a natural conclusion.

A total of five replicates of the biotransformation process were carried out. These five process replicates were subsequently split into two dataset, one for the construction of a calibration model and another for use as an external validation dataset to challenge the constructed model. In order to generate a robust calibration model, the calibration dataset should be representative of the typical variation expected in the biotransformation process.²¹

To ensure that the spectral variations between replicates of the biotransformation were encompassed in the calibration model principal component analysis (PCA) was used, in conjunction with knowledge of the substrates' batch preparations, to classify each process replicate as either a calibration or validation dataset.

All the acquired mid infrared spectra were converted to the second derivative format, and the region of interest ($900\text{-}1620\text{ cm}^{-1}$) extracted. The data was mean centered and each biotransformation process identified as a class for ease of identification. Principal component analysis was carried out on the data, with the constructed model based on two latent variables. The scores plot was examined (plot of PC1 scores vs. PC2 scores) to indicate the variation observed in the spectra (Figure 4-10).

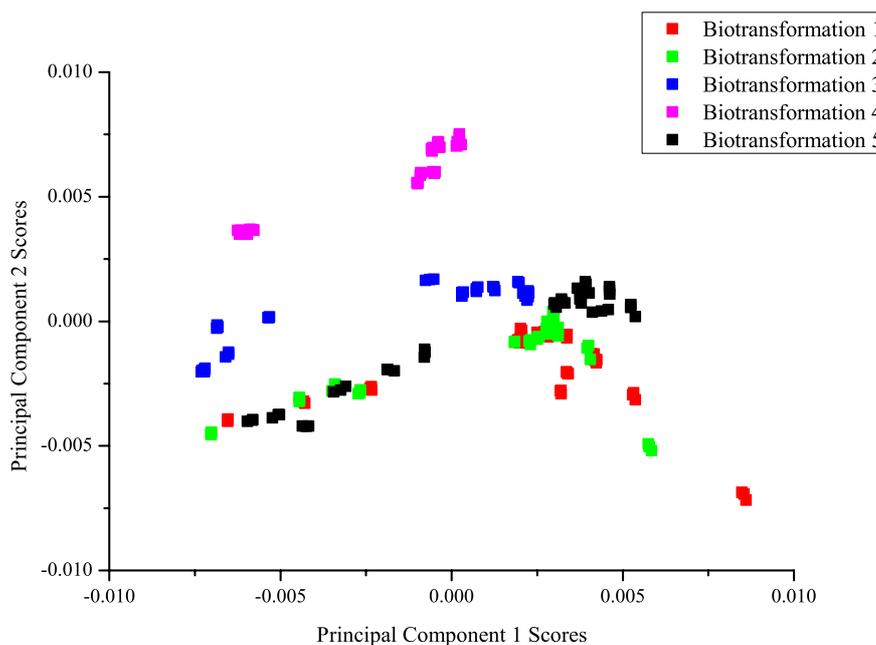


Figure 4-10 - Plot of scores associated with principal component 2 plotted against scores associated with principal component 1 used to identify the most suitable datasets for calibration and the most suitable for validation of the constructed model.

The scores plot indicates that generally there was very good consistency between the five replicates of the biotransformation process. All replicates appear to adopt a similar reaction profile along principal component 1 but do illustrate some differences in the second principal component.

Of substantial note was biotransformation run 4, which appeared to occupy a very different position in the principal component space relative to the other biotransformation replicates. This can be explained by the initial process

concentrations of racemic amino acid that were employed in this biotransformation replicate. Run four had a lower initial concentration due to limitations on the volume of remaining D/L-TBG from the second batch preparation. Conversion of this batch had not progressed as optimally as expected, and so the concentration of racemic TBG was lower and residual TMP higher than in the other batches requiring a higher initial loading volume, and therefore limiting the volume that remained after replicate three had been carried out.

It was clear from this scores plot which replicates were candidates for the calibration dataset so that the model incorporated data that encompassed the variation expected in the process. However before selecting the process replicates for calibration, the PCA scores plot was combined with the knowledge of the batch preparations of racemic amino acid substrate to ensure that the calibration dataset incorporated samples from all three batches prepared.

The batch preparations were utilised in the sequence they were prepared, with each batch being utilised for a maximum of two replicates of the biotransformation process.

A total of three replicates were selected for inclusion in the calibration model, based on the PCA scores plot and the batch preparations biotransformation replicates 2, 4 and 5 were selected. This ensured that the variation observed over the various process replicates was built into the calibration model, but also that the model was representative of the batch nature by which the substrate was generated.

Replicates 1 and 3 were retained as the validation datasets as these were then two independent datasets not included in the calibration model, but were within the predictive ability of the model.

A PLS model was constructed using the spectra obtained for this calibration dataset. The spectrum of each sample was acquired in triplicate, so initially all three spectral replicates were utilised along with the three quantified values obtained for each

sample (from the triplicate injections on the HPLC). Adopting this approach rather than averaging the values ensured that the resulting model took into account variations in the spectra, and also the error associated with the reference method.

In a few cases there were points where the reference analysis had failed to return a result (in most cases this was due to a failure in the derivatisation reaction) or the returned concentrations were inconsistent with the general trend being observed. To prevent these samples from biasing the constructed model the samples were either reanalysed or removed. Aside from the assumption that samples taken were representative of the process, due to the nature of the biotransformation, there was a heavy reliance on the heating step employed being sufficient to denature the enzyme and prevent further conversion of the amino acid outwith the reactor. In most cases this was successful, however this could potentially explain the cases where unexpected values were observed.

A total of fourteen samples were removed as outliers from the collected spectra across the five replicates of the biotransformation process. With these samples removed, the PLS model for the system was constructed using the selected fingerprint region of the spectra and the corresponding HPLC reference data. The second derivative spectra were mean centered, with the model being based on two latent variables (92% variance in spectral data and 95% in reference data) which was determined as optimum from the internal validation procedure.

This constructed model was then challenged using the two process replicates identified as the validation datasets. Both datasets were presented to the model independently and the predicted concentrations from the model compared with the reference values.

Table 9 - Summary of key model parameters for the constructed PLS model for the biotransformation process and the external validation datasets.

	TBG (mMol dm⁻³) [r²]	TMP (mMol dm⁻³) [r²]
RMSEC	27.3	15.6
RMSECV	28.1	16.3
RMSEP	(i) 30.9 [0.87] (ii) 42.9 [0.90]	(i) 18.2 [0.97] (ii) 14.5 [0.99]

In this case the errors associated with the constructed model were high but not unreasonably high (Table 9). Increasing the number of latent variables used in the model caused a reduction in the error of calibration but an increase in the error of cross validation and prediction. This suggested that two latent variables was the optimum, as the number of latent variables increased the model began to over-fit the data and a bias was introduced into the system.

These results suggested that it was possible to quantify the two analytes of interest in the system under investigation. Further improvements to the model were required to try and reduce the error associated with the calibration, cross validation and predictions of the two validation datasets.

Re-examination of the pure component mid infrared region (Figure 4-7) suggested that there were a number of key spectral features that were common to both the analytes of interest. As these features are identical in both analytes, it would be difficult to try and regress these absorbance values against concentration, since this may alter in a non-linear relationship. The spectral regions utilised in the model were refined to remove these common spectral regions and also any other areas that contained what was regarded as redundant spectral information. Closer inspection suggested that only the spectral features between 900-1031 cm⁻¹, 1100-1340 cm⁻¹ and 1440-1620 cm⁻¹ were required (Figure 4-11).

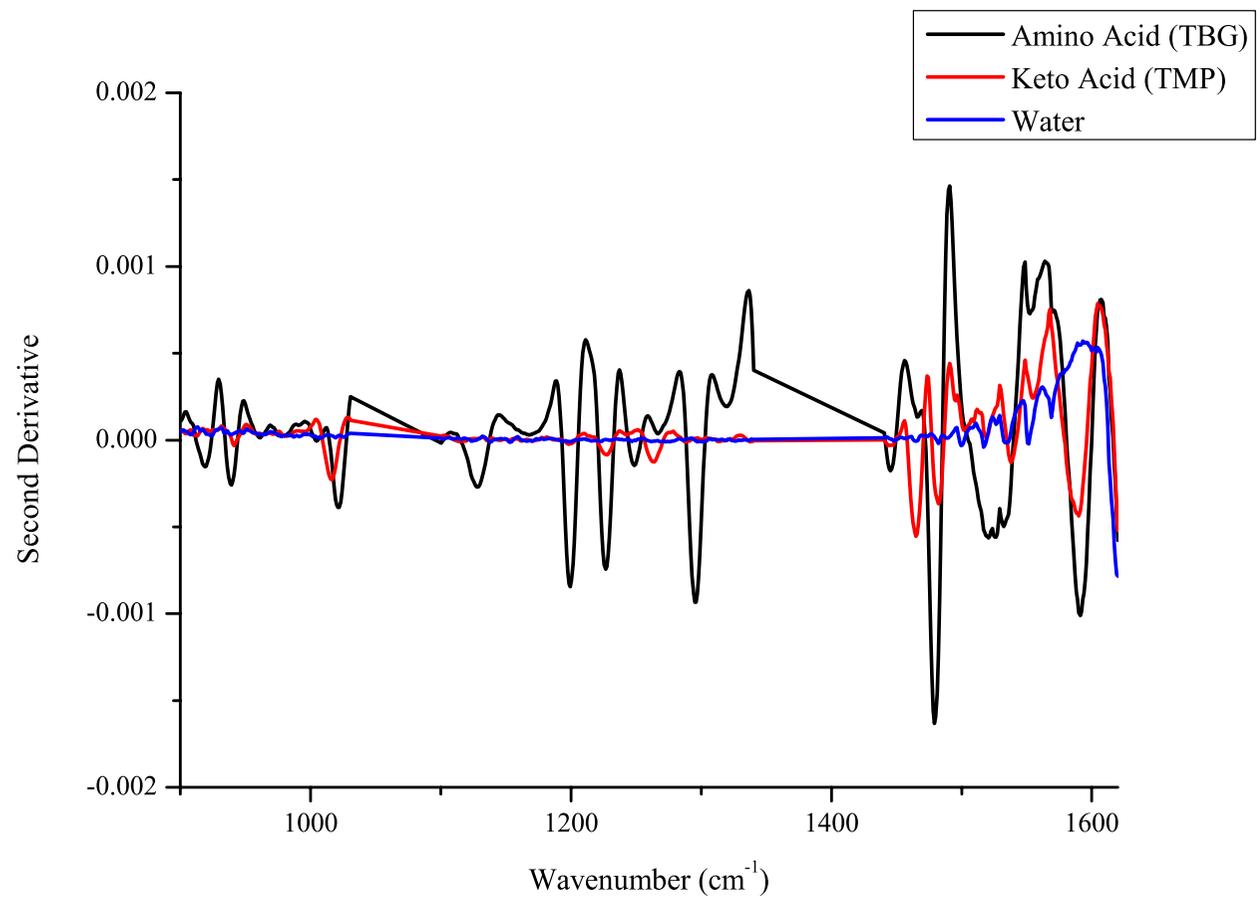


Figure 4-11 – Amended regions removing features identified as being redundant or common to both analytes of interest to improve the model constructed for the system. Regions retained were between 900-1031 cm⁻¹, 1100-1340 cm⁻¹ and 1440-1620 cm⁻¹.

With these clearer defined spectral regions the PLS model for the biotransformation process was reconstructed. As before, the optimum number of latent variables for the system was determined as being two, and the constructed model was challenged using the datasets identified for external validation (process replicates 1 & 3).

Table 10 – Summary of key model parameters and the external validation procedure for the model constructed using the refined spectral regions.

	TBG (mMol dm⁻³) [r²]	TMP (mMol dm⁻³) [r²]
RMSEC	26.7	16.8
RMSECV	27.6	17.5
RMSEP	(i) 33.6 [0.88] (ii) 45.6 [0.87]	(i) 21.1 [0.97] (ii) 13.3 [0.98]

Refinement of the spectral regions resulted in a slight improvement in the error of calibration and cross validation for the amino acid (TBG), but also slightly increased these errors with respect to the keto acid (TMP) (Table 10).

Examination of the scores plot (scores from LV1 against scores from LV2) for the constructed model suggested there was variation or drift being observed within the three replicate spectra acquired of the same sample. Since this was a PLS model some of the drift may be attributable to variations between the replicate injections of the reference analysis. Re-examination of the PCA scores plot (Figure 4-10) confirmed that drift between the replicate scans was also noted, suggesting that some of this variation may also be attributed to drift in the spectra.

The design of the ATR sampling accessory may provide a possible rationale behind these observations. After acquisition of the spectra it was noted that the cell material present in the sample layer had begun to sediment onto the surface of the crystal. This increase in cell material on the surface of the crystal between acquisitions will have an effect on the evanescent wave, and subsequently the acquired spectrum. This was similar to the effect reported by Acha *et al.* (2000) where a hydrophobic coating

was used to exclude water and enhance the spectra.⁵⁷ In this case however the effects were negative with the cell material forming the coating preventing analytes in solution reaching the crystal surface.

Since this sedimentation of the cellular material appeared to be having an adverse effect on the constructed model, the model was re-constructed using only the first of each of the triplicate spectra acquired. The first acquisition of each sample should still be homogeneous and therefore a true representation of the sample matrix under process conditions. As the HPLC analysis was also carried out in triplicate the reference value associated with the spectra was quoted as the mean value of the three replicate injections. Using only the first acquisition and the mean reference values should reduce the error associated with the model and improve the constructed models ability to predict values for the unseen datasets.

Using the whole fingerprint region (900-1620 cm^{-1}) and the mean reference values, a PLS model was constructed as before. Once again the only additional pre-processing applied was to mean centre both the spectral and reference data. LOOCV was applied for internal validation and to determine the optimum number of latent variables, which was again two.

Table 11 – Error values associated with the PLS model constructed from the single spectra of the full fingerprint region and the mean concentration of the reference data.

	TBG (mMol dm^{-3}) [r^2]	TMP (mMol dm^{-3}) [r^2]
RMSEC	25.8	11.9
RMSECV	28.2	13.5
RMSEP	(i) 28.1 [0.89] (ii) 25.7 [0.97]	(i) 18.7 [0.97] (ii) 13.7 [0.99]

The key values associated with this model (Table 11) were overall an improvement over the models that had been previously generated, particularly with regards to the amino acid. Error of calibration values for the amino acid and keto acid were both reduced compared with the previously constructed models. The error of cross

validation for the TBG was fairly consistent with previously observed values with an improvement in the error associated with the TMP component. Where this model demonstrated a substantial improvement over the previously constructed models was in the error of prediction with the unseen validation replicates.

Considering each analyte separately, quantification of TBG in both the validation datasets was in good agreement with the measured values from the reference analysis, with the RMSEP values being close to the RMSEC and RMSECV values. This was a substantial improvement over the previously constructed models where the predicted values varied greatly from the measured values. This was clear graphically from the plot of the predicted TBG (from both validation datasets) against the measured value (obtained from the reference analysis).

With the model constructed using the triplicate spectra and the raw reference data there appeared to be large discrepancies between the measured and predicted values as illustrated by the spread of the points from the central diagonal line (Figure 4-12 (i)). Data points in this plot also do not appear as randomly scattered around the central diagonal line which would suggest a tendency of the model towards over predicting the concentration of TBG. Closer examination of the validation datasets, particularly the triplicate injections carried out to assess repeatability of the method (Table 5) there appears to be a tendency of the derivatisation process to result in an over-prediction of the analyte concentration. Since this data was used in the construction of the model there is the possibility that some of this error has been carried through into the model, and was potentially partially responsible for the tendency towards over predicting the TBG concentrations.

Comparing this with the model constructed using the single spectra and mean reference values (Figure 4-12 (iii)) where the data points appear much closer to the line that represents a 1:1 relationship. Points in this model also appear to be more randomly scattered suggesting there wasn't a bias in the model towards the over prediction of TBG concentration. As well as reducing variations between spectra,

utilising the mean reference values also helped reduce the errors observed between replicate injections resulting in a better calibration model.

In general the errors of calibration, cross validation and prediction with regards to the TMP were not as large as the values obtained for TBG. However this was expected since quantification of this analyte was achieved by direct determination method and not a derivatisation procedure. Using a single spectrum and the mean value for the reference dataset resulted in a reduction in the RMSEC and RMSECV values. However this model did result in a slight increase in the RMSEP values for the two unseen validation datasets. Plots of predicted TMP concentration against measured concentration (Figure 4-12(ii) for triplicate spectra and Figure 4-12(iv) for single spectra) both illustrated data points that were randomly scattered around the central diagonal and did not appear to suggest any bias towards either over or under prediction of the analyte values.

To complete the model iteration process, and attempt to further improve the generated model, a fourth model was constructed utilising the first spectrum of the refined spectral regions (900-1031 cm^{-1} , 1100-1340 cm^{-1} and 1440-1620 cm^{-1}) and the mean reference concentrations. On moving to this model the RMSEC and RMSECV values were consistent with the values obtained in the previous models. Predictions of the concentrations of the two validation datasets however were higher than had been obtained utilising the full fingerprint region (Table 12)

Table 12 – Errors associated with the model constructed using the single spectra of the selected mid infrared regions and the mean values of the reference analysis.

	TBG (mMol dm^{-3}) [r^2]	TMP (mMol dm^{-3}) [r^2]
RMSEC	25.6	12.7
RMSECV	28.3	14.4
RMSEP	(i) 30.5 [0.91] (ii) 26.4 [0.98]	(i) 22.1 [0.97] (ii) 11.3 [0.99]

Based on these results the best model constructed for the biotransformation process under investigation utilised the single spectra of the full fingerprint region and the mean value of the triplicate injections for the reference analysis. This resulted in a model that returned acceptable errors of calibration, cross validation and prediction of the two unseen validation datasets.

Having constructed a suitable model for the process, and tested it using two unseen independent datasets, it was necessary to examine the model in greater detail to ensure that the co-linearity within the system had been addressed and that the model was able to independently quantify both the analytes of interest.

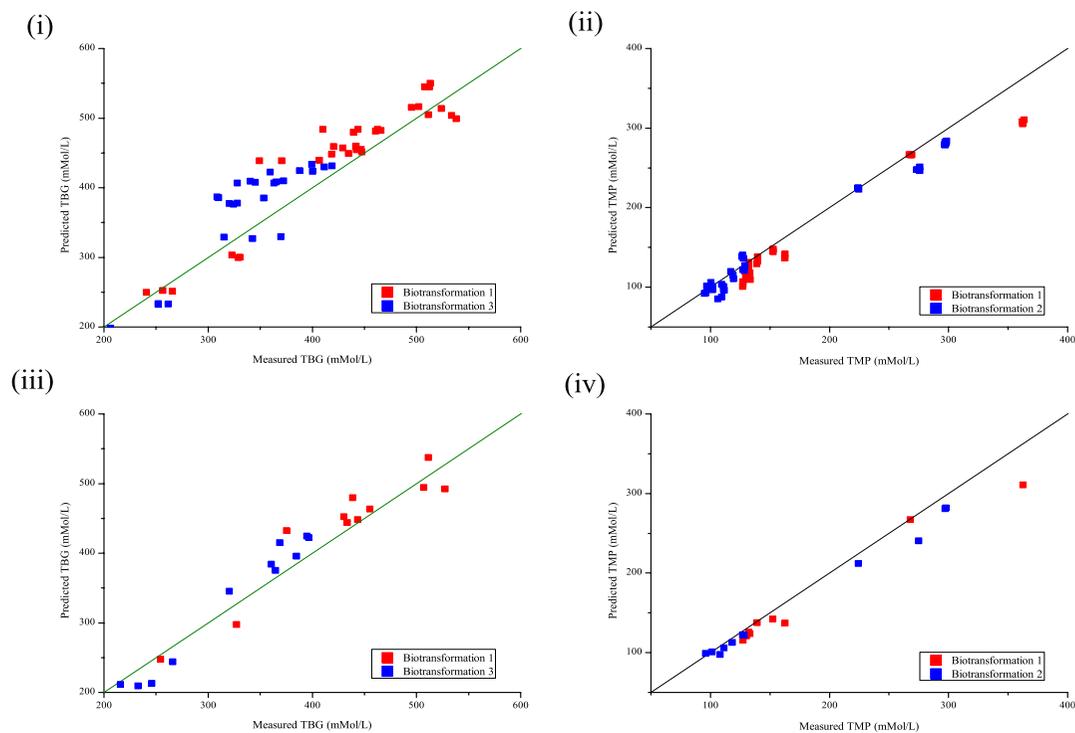


Figure 4-12 – Plots of predicted analyte concentration (mMol dm^{-3}) against measured analyte concentration (mMol dm^{-3}) for the validation biotransformation replicates. (i) and (ii) Analyte concentration predictions for validation datasets based on models constructed using the triplicate spectra and raw HPLC reference data values. (iii) and (iv) Analyte concentration predictions for validation datasets using only the single spectra and the mean reference data value.

4.6 Model Evaluation

Having constructed a PLS model that returned acceptable errors of prediction for the validation datasets, the contributors to the model were examined. Evaluation of the contributors to a constructed model for the system was essential to ensure that the model was monitoring the particular analytes of interest. In such complex biological systems where stoichiometric linkages between analytes were prevalent it was essential to ensure the resulting model was independently quantifying each analyte of interest and not elucidating information from other spectral features or relationships.³⁸

Loadings plots and regression coefficients associated with the model were examined and compared with the pure component spectra to identify exactly what the constructed model was modelling and if co-linearity within the model was an issue.

The loadings associated with the first latent variable (Figure 4-13) had a number of features that were attributable to both the amino acid and keto acid. Signals in the loadings plot that corresponded with the amino acid were almost the mirror image of the signals observed in the pure component spectrum. Since those points that have a high positive loading value correspond with the apex of a trough in the second derivative spectrum this would suggest that a correlated relationship existed between TBG concentration and the first latent variable.

Points in the pure component spectrum of the keto acid corresponded with negative loadings values associated with the first latent variable. This would therefore suggest an anti-correlated relationship between keto acid concentration and the loadings associated with latent variable one.

Of particular note was the absence of any relationship, correlated or anti-correlated, between the amine functional group of the amino acid ($1500-1550\text{ cm}^{-1}$) and the first latent variable.

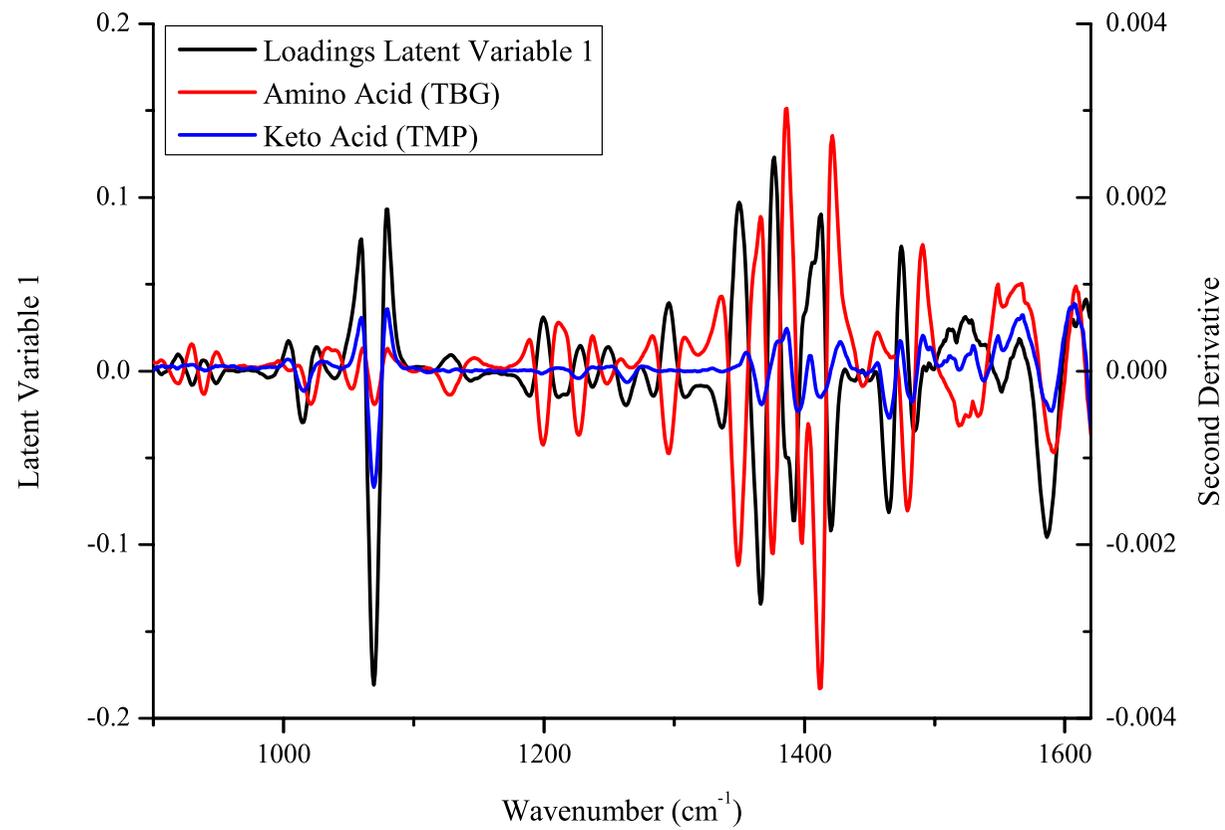


Figure 4-13 - Loadings associated with latent variable one in the constructed PLS model. Overlaid are plots of the pure component amino acid (TBG) and keto acid (TMP) for comparison.

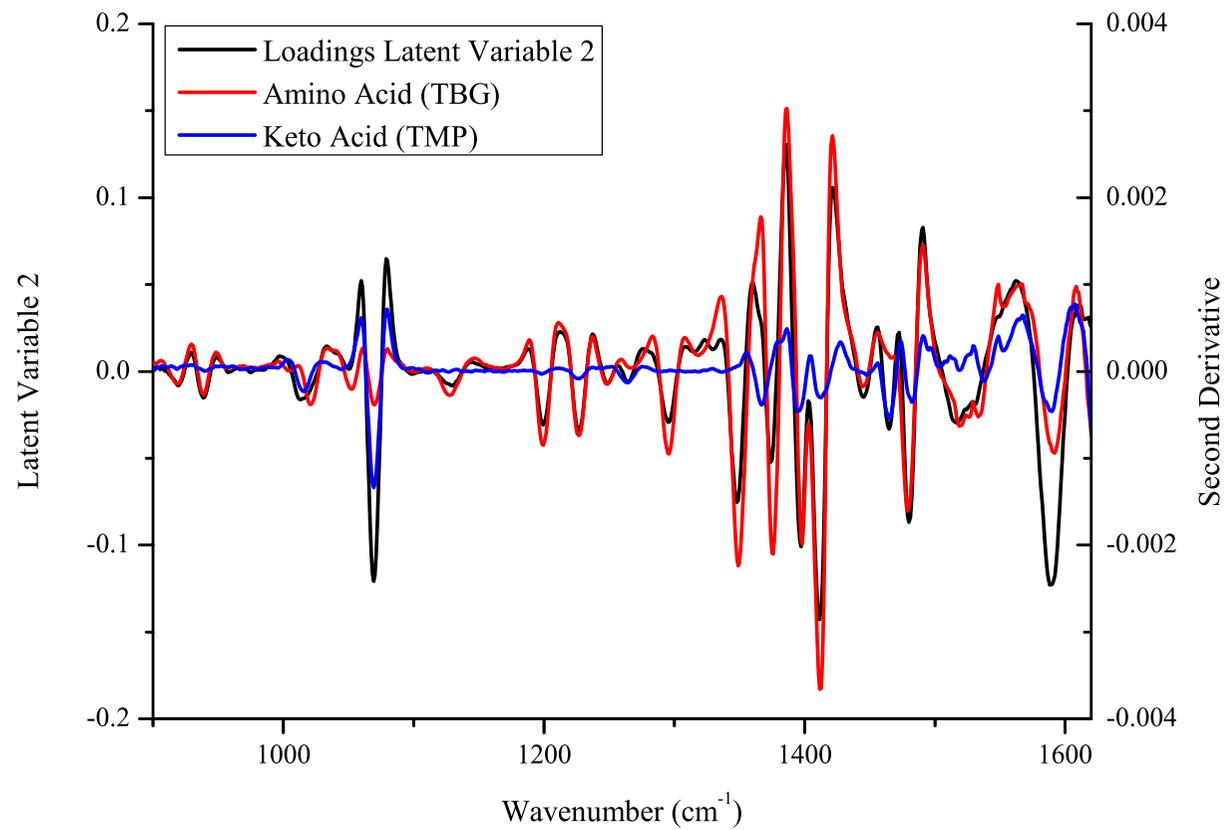


Figure 4-14 - Loadings associated with latent variable two in the constructed PLS model for the biotransformation process. Again overlaid are the pure component spectra of the amino acid (TBG) and keto acid (TMP) for comparison.

Given that the two analytes of interest were chemically very similar, the vast majority of variations observed in the spectra would have been due to features common to both analytes. Also given the stoichiometric link that exists between the two compounds it was not unsurprising that the first latent variable contained spectral features attributable to both the key analytes since these features were the main source of variation within the spectra.

Comparison of the loadings associated with the second latent variable (Figure 4-14) did indicate some association with the amine signals of the amino acid. These were in addition to a number of other points where negative loading values corresponded with troughs in the pure component spectra of the amino acid, therefore suggesting an anti-correlated relationship. The second latent variable did not appear to show any correlation or anti-correlation with spectral features associated with the keto acid (TMP).

Evaluation of the loadings associated with the two latent variables used in the construction of the model suggested each analyte was being independently quantified. The absence of the amine signals from the first latent variable, and their presence in the second, indicated changes in the amine spectral features were being identified as variations that were not linked with the bulk change in the system described by the first latent variable.

To confirm that the analytes of interest were being independently modelled by the system, thus addressing the issue of co-linearity in the system due to stoichiometric linkages, the regression coefficients associated with the constructed model were also examined.

The first regression coefficient for the model appeared to be an exact match with the pure component spectrum of the amino acid. Points with a non-zero regression coefficient value corresponded with the location of signals observed in the second derivative pure component spectrum of the amino acid. Of particular note is the non-

zero regression coefficient value that corresponds with the amine region (1500-1550 cm^{-1}).

The second regression coefficient associated with the constructed model was also examined along with the pure component amino and keto acid. In this case the non-zero regression coefficient values appeared to correspond with the signals observed in the pure component keto acid spectrum. Focusing on the region where signals from the amine functional group were noted (1500-1550 cm^{-1}), this regression coefficient has a zero value in this region. Near zero values for this regression coefficient were also observed at other points where unique spectral features of the amino acid were noted.

Should the constructed model be inferring the concentration of a particular analyte from the spectral features of another, this would be reflected in the regression coefficients. The fact that both regression coefficients had unique features that could be correlated with spectral features associated with either TBG and TMP suggested that the model was able to independently quantify both the analytes of interest. Had co-linearity been an issue in the system both regression coefficients would be expected to have a number of common features that could be attributed to both the analytes of interest.

Given that this was a biological system with a known correlation between the two analytes of interest, it was important that the constructed model was sufficiently robust to allow independent quantification of the analytes. A number of approaches to addressing this issue of co-linearity have been extensively discussed. Commonly methods involving the use of synthetic samples, the spiking of media samples or the retrospective adulteration of actual samples have been employed.¹⁷

Evaluation of the constructed model would suggest that co-linearity within the model had been suitably addressed, however none of the aforementioned spiking processes were employed. With this particular case the process by which the racemic amino

acid substrate was prepared (3.3.2.1) and how the datasets were selected (4.5) provided sufficient variation to allow construction of a robust model.

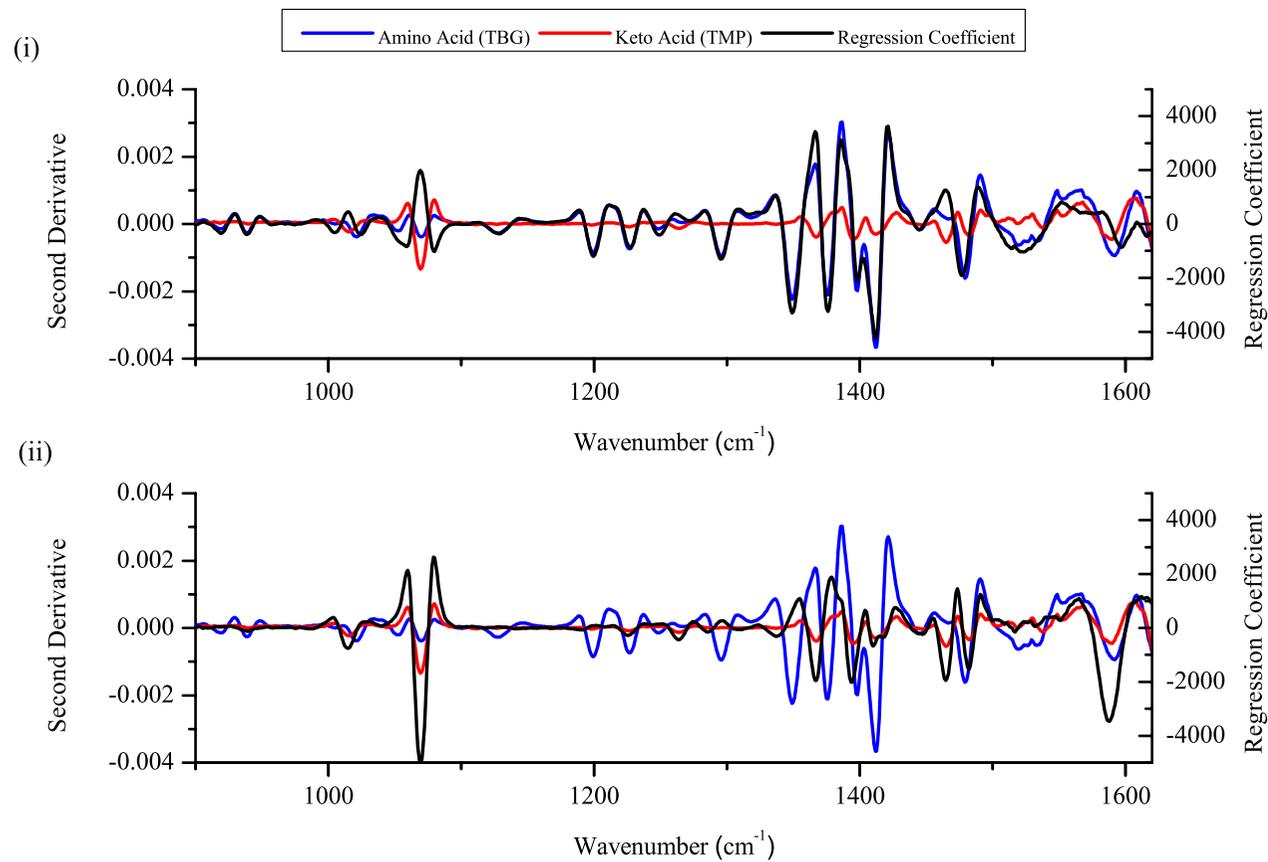


Figure 4-15 - Regression coefficients associated with the constructed PLS model plotted alongside the pure component spectra of the amino acid (TBG) and keto acid (TMP) for comparison. (i) Plots the first regression coefficient against the pure component spectra. (ii) Plots the second regression coefficient with the pure component spectra.

The batch nature of the reductive amination stage, converting TMP into racemic TBG, meant each batch contained a slightly different final TBG concentration and residual TMP concentration due to factors such as the initial loading concentration and catalyst efficiency. Selection of a process replicate from each batch for inclusion in the calibration model also helped ensure that the concentrations of both analytes were independent from the outset, even though as the process progressed the concentrations varied in a stoichiometric manner.

Ehly *et al* (2007) reported adopting a similar approach to dealing with co-linearity within a multi-batch system.¹³⁶ Alterations to the initial ratio of substrate were carried out to help break the co-linearity within the system. Preparation of the amino acid substrate in this case by the batch reductive amination process, and the inclusion of a replicate from each batch in the calibration model, ensured that the initial concentrations of TBG and TMP were independent of each other at the start of each biotransformation process replicate.

Adopting the approach described by Ehly *et al.* (2007) would remove the requirement for retrospective sample adulteration necessary in the more traditional spiking approaches to addressing co-linearity within the system. Tamburini *et al.* (2003) and Arnold *et al.* (2003) both defend the use of actual process samples in the model construction process, arguing that the interactions between the various components of the sample matrix vary from batch to batch and that by using these real samples the resulting model will be much more robust compared with a model constructed using synthetic spectra.^{48, 49} The approach of varying the initial concentrations of substrate allowed the co-linearity issues to be addressed whilst still utilising actual process samples to construct the model, so any side reactions or interactions were still being incorporated in the model when adopting this approach.

In this particular case examination of the contributors to the model indicated that the co-linearity issue had been suitably dealt with, greater confidence in the models ability to independently predict analyte concentration could be obtained by adopting

a more structured and analytical based approach towards varying the initial ratios of substrate, and product if required.

4.7 Model Application

The constructed model was externally validated using two datasets that were independent from the calibration dataset. To further validate and challenge the constructed model an additional two replicates of the biotransformation process were carried out. These replicates were carried out in a different location and in a completely different reactor system (Ingenza in-house system) compared with the replicates used in the calibration and validation stages.

Both off-site replicates of the biotransformation process were carried out utilising a single batch preparation of the racemic amino acid (TBG). In this case the design of the reactor vessel was substantially different from the bioreactor system that had been utilised to carry out the process replicates used for calibration and validation. The reactor vessel in which these processes were carried out was a basic heated water jacket vessel with agitation and aeration provided via a rudimentary setup. Since an at-line sampling approach was applied the design of the reactor vessel was not expected to have a major influence on the quality of the spectra or the predictive ability of the model however there was some expected differences in the behaviour of the process that needed consideration.

The Applikon bioreactor system employed for the calibration and validation replicates was a basic stirred tank reactor (STR) design. Agitation in the system was provided via two Rushton design impellers with baffles fitted in the vessel to prevent vortex mixing thus also enhancing oxygen transfer.¹⁴ Air was supplied into the system by means of a ring sparger situated directly under the first impeller. The whole design of this system was intended with the purpose of maximising oxygen transfer in the media.^{14, 137} Higher oxygen levels in the media would therefore allow faster regeneration of the FAD energy source required by the DAAO enzyme and therefore allow the biotransformation to progress at a faster rate. In addition the

bioreactor based system allowed for measurement and control of important parameters such as dissolved oxygen, pH and temperature within a single system.

With the in-house heated water jacket vessel system, agitation was provided via a single marine design impeller and in this case the reactor vessel was not equipped with baffles to assist in the mixing process. Sparger design in this system was also substantially different from the bioreactor system, with air introduced via a dip pipe design sparger located under the impeller.^{14, 137} Combined all these features were likely to result in an overall reduction of dissolved oxygen levels in the process medium, therefore reducing the rate at which the process occurred by limiting the oxygen available for regeneration of the FAD energy source required by the DAAO enzyme. In common with the bioreactor system dissolved oxygen and pH measurements were made using the relevant probe systems however the measurements were not linked or controlled in a manner comparable with the Applikon bioreactor system.

The two different reactor system designs were unlikely to have any substantial impact on the ability of the constructed model to predict the key analyte concentrations. Although the samples were measured at-line it could potentially be argued that the better oxygen transfer in the bioreactor system resulted in high levels of dissolved gases, which coupled with the impeller design and baffles would result in an increase in the number but decrease in size of the air bubbles which may cause a change to the acquired spectrum. The differences were however, more likely to influence the rate at which the reaction progressed and so initially the changes observed in the spectra would be relatively minor which may affect the accuracy of the predictions.

A total of two additional replicates of the biotransformation process were carried out adopting the same procedure as described (3.3.2) and the mid infrared spectra of the samples acquired.

As expected the biotransformation processed did not progress as rapidly as had been found in the bioreactor based systems, as indicated by a slower rate of reaction and smaller overall conversion of amino acid in the same time frame.

The acquired spectra underwent the same pre-treatment stages and region selection processes as had been previously utilised.

These spectra were then presented to the constructed model and the concentration of the two key analytes predicted and compared with the reference values obtained using the off-line HPLC method.

On this occasion the model did not appear to be predicting the concentration of the two key analytes particularly well (Table 13). For both replicates of the process the RMSEP values obtained were substantially higher than had been observed with the external validation datasets.

Table 13 – Error values associated with the prediction of the two off-site biotransformation replicates.

	Amino Acid (TBG) (mMol dm⁻³) [r²]	Keto Acid (TMP) (mMol dm⁻³) [r²]
Model RMSEC	25.8	11.9
Model RMSECV	28.2	13.5
Replicate 1 RMSEP	44.9 [0.81]	46.5 [0.80]
Replicate 2 RMSEP	42.4 [0.99]	41.5 [0.99]

Examination of the plots of predicted concentration, as determined by the constructed model, against the measured concentration from the reference analysis confirmed that the model was failing to correctly quantify the key analyte concentrations (Figure 4-16)

These plots clearly illustrate a tendency of the model towards an over prediction of analyte concentration by the model. This was the case for both the analytes of interest but was particularly bad for the keto acid (TMP) component.

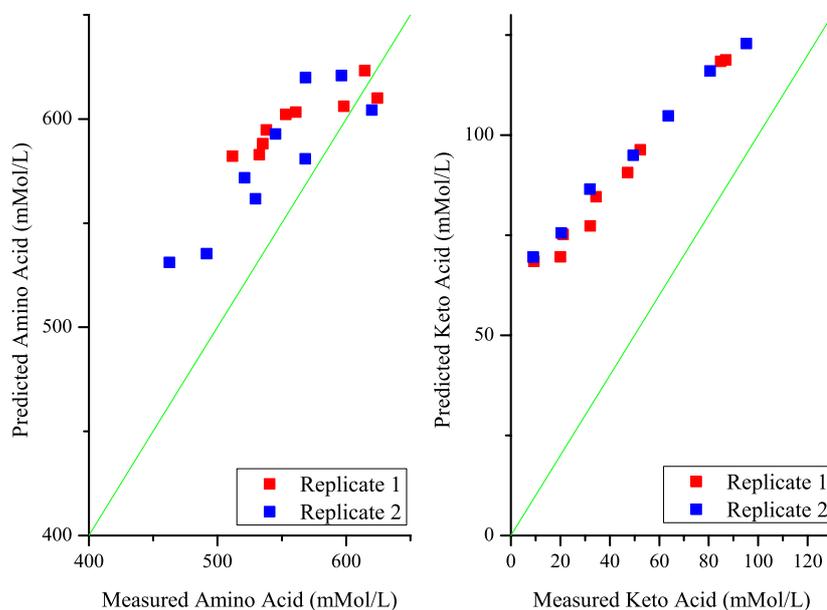


Figure 4-16 – Predicted concentration (mMol dm^{-3}) plotted against measured concentration (mMol dm^{-3}) of both amino acid and keto acid for biotransformation processes carried out at different site from those used to construct and validate the PLS model.

This failure of the model to deal with the biotransformation process carried out at a different location was surprising. As this was a mid infrared model, constructed using the fingerprint region, the signals that arose were attributed to fundamental vibrations of the molecules. Compounds utilised in the calibration and validation datasets were identical to those utilised in these additional replicates of the process and would therefore be expected to generate the same infrared spectrum.

Variations in the spectrometer system utilised for the acquisition of the spectra or the mechanism by which the sample was presented have been reported to influence the quality of the spectra acquired and hence the need for maintenance of the model.¹²²

However, in the current situation there was no alteration to the spectrometer or the sample presentation mechanism – spectral acquisition was carried out at-line using the same instrument and ATR crystal used to acquire the spectra for the calibration datasets.

A potential hypothesis as to why the model was failing to accurately predict the concentration of these key analytes relates to the dissolved oxygen levels in the system. There are many reported cases of high levels of dissolved gases and the highly light scattering nature of the matrix causing challenges for modelling in the near infrared region.^{26, 32, 46} Whilst theoretically the ATR technology utilised with the mid infrared region should reduce these effects, due to the much shorter path lengths, there are some suggestions they could potentially be problematic.^{51, 55}

Dissolved oxygen has been reported to have an effect on the refractive index value of water.¹³⁸ Given that the biotransformation process was carried out in aqueous media the differences between the two reactor systems may have resulted in a difference in the refractive index of the sample material. Acquisition of the infrared spectra using the ATR relies on the refractive index of the sample material being less than the refractive index of the crystal material.⁵⁸ Changes in the refractive index of a sample material have been reported to cause shifts in both the absorption intensity and peak position within the UV-VIS region.¹³⁹ It could therefore be possible that a similar phenomenon was being observed in this case as a result of the different reactor vessel systems utilised.

A selection of spectra obtained from the calibration dataset was compared with the spectra obtained from the replicate carried out at the different location. Visually there did appear to be some slight shifts in the spectra although these were very minor. Given that this was based on a multivariate model, PCA was used with only the spectral data to determine if any differences in the spectra could be identified.

From the PCA scores plot (Figure 4-17) it was clear that the spectra of the samples from the biotransformation process carried out in the heated water jacket vessel

occupied a different position in the principal component space relative to the calibration datasets. There were however some similarities in the spectra as evident by their similarities to the calibration dataset in relation to the second principal component.

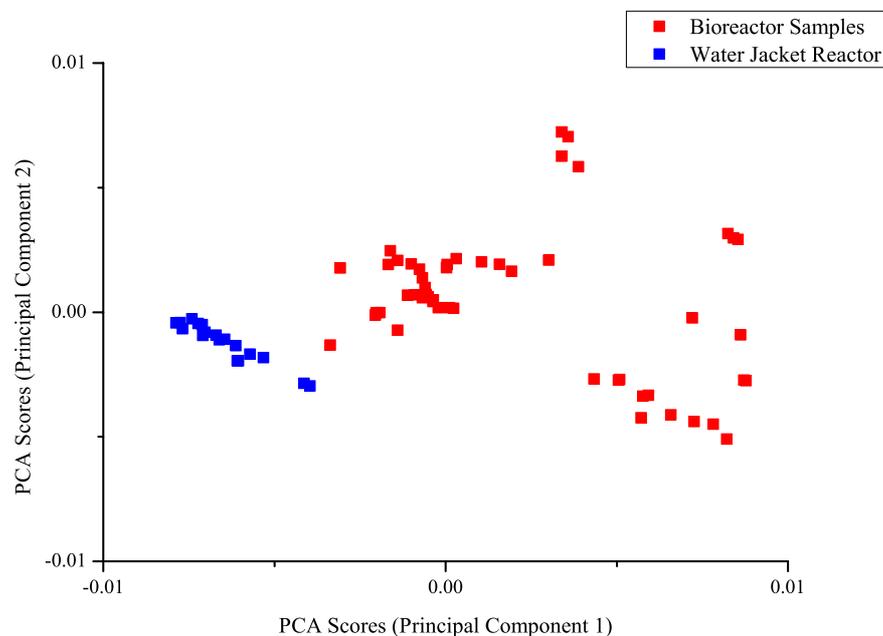


Figure 4-17 - Principal component analysis carried out on samples used in the construction of the calibration model (that were carried out in the Applikon bioreactor system) and those carried out off-site in the heated water jacket vessel.

Effects on the acquired spectra as a result of the differences in dissolved gas content of the samples may offer some explanation as to why the model failed to accurately predict the analyte concentrations.

Further investigation would be required to confirm if this were the case. The most logical approach would be to look at a mechanism of degassing the samples prior to acquiring their infrared spectrum. Removal of the gases should ensure that the refractive index of the samples from both systems was consistent and the spectra resulting from both were comparable.

A second possible explanation for the failure of the model could be attributed to the raw materials used in the biotransformation process. In the case of the replicates carried out off site, the TMP starting material used in the reductive amination to generate the racemic TBG was purchased from a different supplier to that used to generate the racemic material for the calibration dataset. As previously discussed this model was based on the mid infrared spectra, the signals arose due to fundamental vibrations of bonds in the molecule so the observed spectrum should have been consistent regardless of supplier. There may however be some spectral features or shifts that result from impurities that may be present in one manufacturers material not found in the others.

Since infrared spectroscopy can be utilised as a screening method to assess substrate batch quality for processes⁷⁵, it would be a logical progression to suggest that the variability in the batches from different manufacturers were impacting on the ability of the constructed model to predict the key analyte concentrations.

External validation of the constructed model demonstrated its ability to predict the key analyte concentrations from the acquired spectra within acceptable error margins. Further challenging the model with the off-site replicates, and its subsequent failure to accurately predict the analyte concentrations suggested there was insufficient variance in the calibration dataset. Variation arising from the utilising different substrates may be incorporated into the model by including some process replicates using these batches into the calibration model. This may improve the predictions from the model however it would require constant maintenance. Each time there was a change in supplier of raw materials the model would require updating, an expensive as well as time and labour intensive process.

Should the variance and poor predictions arise as a result of the different reactor systems, then addressing these issues and improving the quality of the predictions may be a more difficult process. Incorporating these variances into the calibration dataset would likely result in higher errors of calibration, meaning the predicted concentration of analytes in samples from either system would not be as accurate.

These results highlight the need for process consistency when attempting to apply infrared spectroscopy as a mechanism of process analysis. Deviations in the equipment, instrumentation or substrates can have a substantial effect on the accuracy of the predicted concentrations of the compounds of interest.

4.8 Summary & Conclusions

Using the fingerprint region of the mid infrared region, a PLS model was constructed for the biotransformation process where DAAO enzymes were utilised to selectively convert the D-amino acid in a racemic mixture to the corresponding keto acid. The model was constructed based on a modest number of process replicates (three) and validated using a further two process replicates that were independent of the calibration dataset.

Examination of the contributors to the resulting model suggested that any co-linearity within the system had been addressed and that the resulting model was able to independently quantify both the analytes of interest.

The predicted concentrations of the validation replicates from the model were in reasonable agreement with the values determined by the reference analysis methods (Figure 4-18). Prediction of the concentration of KBA was generally more accurate than predictions of the TBG concentration, however this was not unsurprising given quantification of TMP was via a direct determination method whereas the TBG required a pre-column derivatisation stage.

Replicates of the biotransformation process carried out off-site were also presented to the constructed model and the ability of the model to predict the analyte concentration in these samples assessed. The model was unable to predict these concentrations at a satisfactory level, with both the amino acid and keto acid predictions varying substantially from the concentrations determined by the off-line reference analysis methods. Physical phenomena and variations in substrate production processes were postulated as possible explanations for these failures.

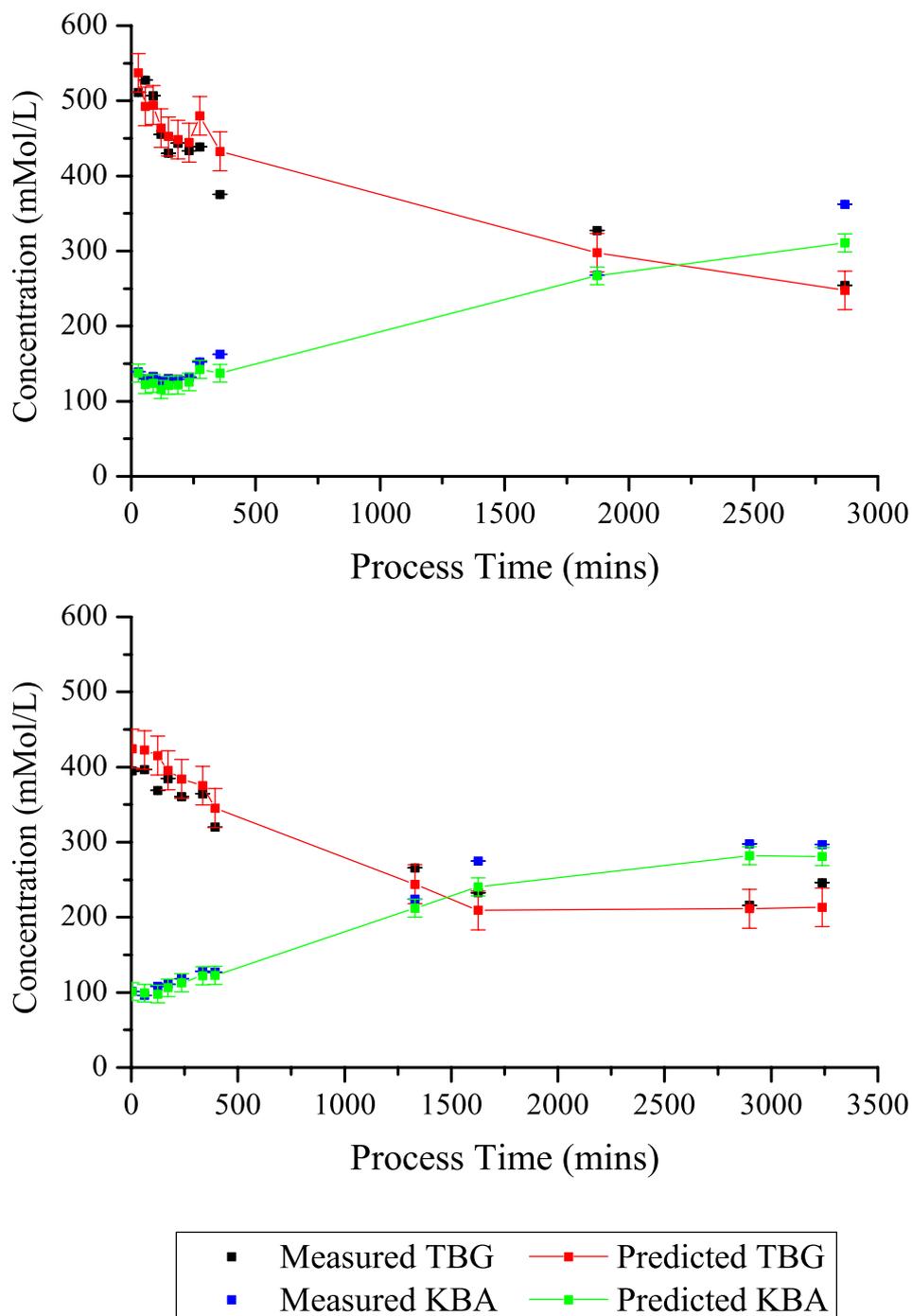


Figure 4-18 - Summary plots of concentrations of TBG and TMP determined from the reference method (HPLC) and the PLS model for (i) validation set 1 and (ii) validation set 2. HPLC values for TBG are quoted as $\pm 0.27 \text{ mMol dm}^{-3}$ and TMP as $\pm 0.14 \text{ mMol dm}^{-3}$ as calculated. Concentrations predicted from model were quoted as $\pm 25.8 \text{ mMol dm}^{-3}$ and $\pm 11.9 \text{ mMol dm}^{-3}$ respectively.

These findings suggested that it was feasible to utilise the fingerprint region of the mid infrared spectrum to construct a quantitative model for this particular industrial biotransformation process. The resulting model was adequate but the process highlighted a number of areas where improvements could be made, and issues that required to be addressed, when constructing future models for similar biotransformation processes.

5. Small Scale Biotransformation

5.1 Process Overview

Following on from the amino acid de-racemisation method where the D-enantiomer of TBG was selectively converted to TMP using DAAO enzymes contained within an *E. coli* host cell (4), other similar biotransformation processes of industrial relevance were considered.

The second biotransformation process to be investigated was once again the selective de-racemisation of a racemic amino acid mixture using the D-amino acid oxidase enzyme. In this case however, the system differed slightly in that the host organism was *Pichia pastoris* modified to express the DAAO enzyme, not *E. coli* as had been utilised previously. The product of interest was the L-enantiomer of amino butyric acid (ABA), with the D-enantiomer undergoing conversion to the corresponding keto butyric acid (KBA) (Figure 5-1) as detailed in section 3.3.3.

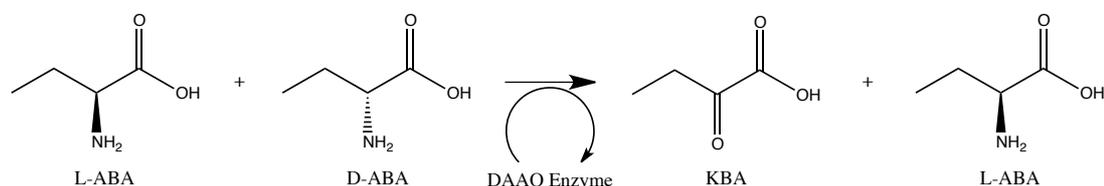


Figure 5-1 – Schematic representation of the biotransformation of D-amino acid (ABA) to the keto acid (KBA) using the DAAO enzyme contained within a *Pichia pastoris* host.

As with the previous biotransformation, the rationale behind this work stems from the limited process information that is available in near real time for this particular class of bioprocess. More near real time information will allow for much greater understanding, control and optimisation of the process, as well as being aligned with the aims of the FDAs PAT initiative.⁶

5.2 Aim & Objective

The aims and objectives in this work was to investigate the feasibility of applying near and mid infrared spectroscopy to monitor an industrial biotransformation process. Although the TBG biotransformation process (Chapter 4) had suggested the near infrared region was unsuitable for this purpose, the different analytes used in this case warranted further investigation of this region.

Once the suitable regions of the infrared spectrum were identified, this data was utilised to construct and externally validate a quantitative model for the two key analytes of interest (ABA and KBA). This model was also to be challenged using a process replicate carried out off-site.

With this particular biotransformation, the reactants were purchased directly from commercial suppliers and not produced by a batch process carried out in the laboratory. This resulted in an additional challenge of ensuring that co-linearity within the system was adequately addressed to allow independent quantification of the key analytes by the model despite a known stoichiometric linkage. Contributors to the model should also be examined to ensure this was the case.

5.2.1 Novelty

As with the other biotransformation systems investigated, these systems have not previously been extensively investigated using either near or mid infrared spectroscopy. The particular biotransformation process under investigation, the selective de-racemisation of a chiral amino acid mixture, provides an additional level of complexity and novelty within the system.

Within this chapter the mechanisms employed to break the co-linearity within the system represent a novel approach. Stoichiometric relationships within the system have been a key point of discussion. Some argue that the relationships should be incorporated into the model⁴⁸, others take a retrospective approach to altering the

samples to break this co-linearity.^{17,36} However there doesn't appear to be a method where the concentrations of key analytes are varied by design to not only break co-linearity but also maintain the sample matrix. The proposed approach varied the analyte concentrations from the outset according to an experimental design, breaking co-linearity from the outset by adopting a structured approach. In dealing with co-linearity in this manner there was the potential to construct an *in-situ* model that also adequately dealt with co-linearity.

5.3 Analytical Reference Method

Development and validation of analytical methods for the quantification of the key analytes of interest was required prior to carrying out any replicates of the actual biotransformation process.

5.3.1 Amino Butyric Acid

5.3.1.1 Method Development

Quantification of the ABA content in the samples was achieved based on adaptations of the analytical method employed for the separation and quantification of TBG (3.2.1.2.1). Pre-column derivatisation of the amino acid was again achieved using a mixture of mercaptoethanol and OPA. The gradient elution method developed for the quantification of TBG was used as a starting point. This gradient program was altered until an acceptable peak shape was obtained with good baseline separation between the peak of interest and any other components in the sample. The conditions described (3.2.1.2.2) represent the optimum gradient elution program for the separation and quantification of the amino acid component.

Using these conditions, the resulting peak returned an asymmetry factor of 1 and a peak tailing factor of 1.03. These values confirmed that the resulting peak was symmetrical and that there was no obvious tailing or fronting observed which might affect the integration parameters.

5.3.1.2 Method Validation

In line with the previously adopted approach (4.3.1.2), a series of ABA solutions were prepared at varying concentrations over the range of interest and analysed using the developed method to ensure a linear response was obtained and to estimate the error associated with this method.

A total of six solutions at varying ABA concentrations between 2 mMol dm⁻³ and 12 mMol dm⁻³ were prepared and analysed to construct a calibration curve, with each sample being analysed in triplicate. Peak area response was plotted against concentration and a linear line of best fit plotted through the data points (Figure 5-2).

Using this equation, the error attributed to fitting a linear line through the data points, i.e. the lack of fit in the calibration equation, was determined. The predicted concentration of the various samples was compared with the true value and the root mean square error value determined as being ± 0.18 mMol dm⁻³ (Appendix II).

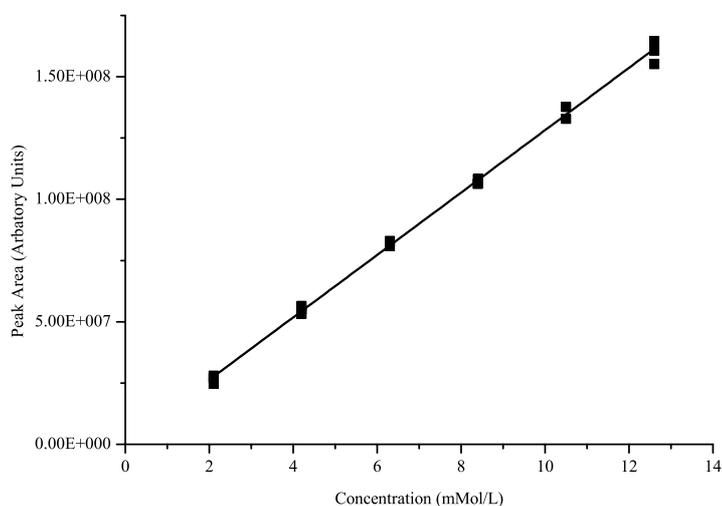


Figure 5-2 - Calibration curve for quantification of ABA showing linear line of best fit through the data.

Ten replicate injections of the same sample were also carried out to estimate the error that arose due to variations in replicate injections. Since this method was not a direct determination of the amino acid concentration, estimation of this error component in the overall error was particularly important since it gave an indication of the reproducibility of the pre-column derivatisation process. The error associated with these replicate injections was determined as being $\pm 0.06 \text{ mMol dm}^{-3}$ (Appendix II). Given the pre-column derivatisation stage this error was very good indicating a consistent derivatisation process.

Overall the error associated with the method was calculated as a linear combination of both sources of error, therefore the error in quantification of ABA using this method was determined as being $\pm 0.19 \text{ mMol dm}^{-3}$, marginally better than had been observed with TBG (4.3.1.2).

5.3.2 Keto Butyric Acid

5.3.2.1 Method Development

Attempts to quantify the amount of keto acid present in the samples using a method similar to that employed with the TMP (discussed previously in section 3.2.2.1) failed. KBA does not possess a sufficient chromophore to allow detection using the DAD detector, so it was necessary to utilise a refractive index detector in this case. Also, there did not appear to be sufficient retention of the analyte of interest by the C_{18} column utilised. KBA passed through the chromatography column eluting completely un-retained by the column. It was clear at this point that an alternative column packing material was required to get sufficient column-analyte interactions to achieve the desired separation.

A chromatography column designed for the separation of organic acids based on an ion exchange principle was thus selected for the analysis. The use of a refractive index detector system restricted the employed method to an isocratic mobile phase composition and constant flow. Initially the conditions employed were a replicate of

those suggested by the column manufacturer (0.005 N H₂SO₄ at a flow rate of 1 mL/min). With this new column packing material there was a substantial improvement in analyte retention. Development of the method conditions from this starting point suggested that the optimum conditions for the separation of KBA employed a mobile phase composed of 0.01 N H₂SO₄ at a flow rate of 0.5 mL/min and a column temperature of 70 °C. Under these conditions the analyte had a retention time in the region of 4.5 minutes, and the resulting peak an asymmetry factor of 1 and tailing factor of 1.04, again suggesting a symmetrical peak with no tailing (Appendix II).

5.3.2.2 Method Validation

A series of calibration standards at varying concentrations were analysed using the developed method. A linear line of best fit was plotted through these data points and the equation determined (Figure 5-3). The root mean square error and the error associated with replicate injections were then determined based on this calibration equation (Appendix II).

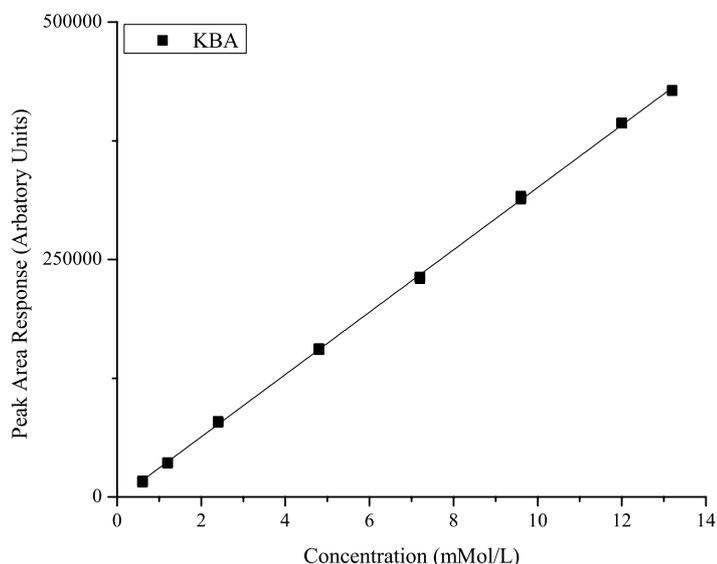


Figure 5-3 – Calibration curve for the quantification of KBA using the developed HPLC method.

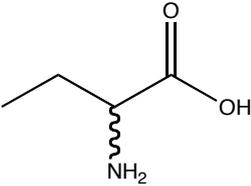
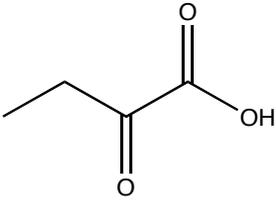
The error associated with the lack of fit of the linear line through the calibration data was determined as being $\pm 0.07 \text{ mMol dm}^{-3}$ and the error between replicate injections as $\pm 0.01 \text{ mMol dm}^{-3}$ giving a total error of ± 0.07 associated with the quantification of KBA for this method (Appendix II), this represented a slightly lower error margin than had been observed for the quantification of TMP (4.3.2.2).

5.4 Initial Feasibility Study

Initially the near infrared region was considered as a potential method for monitoring this biotransformation process. Pure component samples of both the amino acid (ABA) and keto acid (KBA) were prepared in an aqueous background matrix at concentrations representative of those expected during the biotransformation process. The infrared spectra were acquired at-line (3.5.1) before undergoing manipulation and examination to identify regions that exhibited differences, or shifts in the observed signals that could be utilised to construct a multivariate model for the biotransformation process.

The theoretical location of signals arising due to the functional groups present in ABA and KBA were identified (Table 14). These analytes were similar to TBG/TMP with respect to the functional groups present in the molecule. Based on the theoretical locations of signals arising from these functional groups, there did not appear to be substantial variation between the two molecules and so many of the difficulties previous encountered with the near infrared region were likely to recur.

Table 14 - Theoretical locations of signals expected to be observed for ABA and KBA.

Amino Butyric Acid (ABA)		Keto Butyric Acid (KBA)	
			
Functional Group	Regions	Functional Group	Regions
CH ₃	4167-4545 cm ⁻¹ 5882-6060 cm ⁻¹	CH ₃	4167-4545 cm ⁻¹ 5882-6060 cm ⁻¹
CH ₂	5620-5950 cm ⁻¹ 4160-4444 cm ⁻¹	CH ₂	5620-5950 cm ⁻¹ 4160-4444 cm ⁻¹
CH	4081-4385 cm ⁻¹ 5617-5952 cm ⁻¹	COOH	5235-5319 cm ⁻¹
NH ₂	4545-4761 cm ⁻¹	CO	4920-5260 cm ⁻¹ 4360-4690 cm ⁻¹
COOH	5235-5319 cm ⁻¹		

The acquired spectra were converted to the second derivative format, to enhance spectral features and reduce baseline shift, and split into the three spectral regions of interest (second overtones, first overtones and combination bands) to identify features of interest (Figure 5-4).

In the second overtones region, very few signals of interest for the analytes were identified. Due to the weak nature of these signals, the infrequency with which they occur, and close similarities between the analyte molecules, a lack of substantial differences between signals in this region was not unsurprising.

Within the first overtones region some signals that could be attributed to the analytes of interest were beginning to be observed in the 5800-6000 cm⁻¹ region. These signals were most likely due to vibrations of the aliphatic CH bonds. A very subtle difference was noted between the peak maxima of the two analytes of interest. This

difference could potentially be useful for constructing a multivariate model for the system. The signals expected from the amine functional group were not observed, most likely being obscured as a result of the dominant signals from the OH vibrations of the water background matrix.^{24, 97, 140}

In the combination bands region there were a number of signals that could be identified for each of the analytes of interest. Signals between 4250-4450 cm^{-1} were noted for both analytes again likely due to the aliphatic CH vibrations. Like the signals in the first overtones region there also appeared to be a subtle shift in the position of the peaks. A signal that appeared unique to the amino acid component was observed at 4250 cm^{-1} expected to be due the CH vibration at the C_2 position, which is not present in the keto acid.

Although the use of an *in-situ* near infrared probe prevents access to the combination bands region due to absorption of this region by the silica fibres,¹⁴¹ the differences observed in the spectra suggested that it might be feasible to construct a multivariate model for this process using the combination bands region. An at-line model would still represent an improvement over the current off-line analysis methods. The presence of signals in the first overtones region however does suggest that *in-situ* monitoring of the biotransformation process using near infrared spectroscopy may also be feasible.

Since there was a fair degree of similarity between the near infrared spectra of the two analytes of interest, the pure component spectra of the two analytes were also obtained for the mid infrared region.

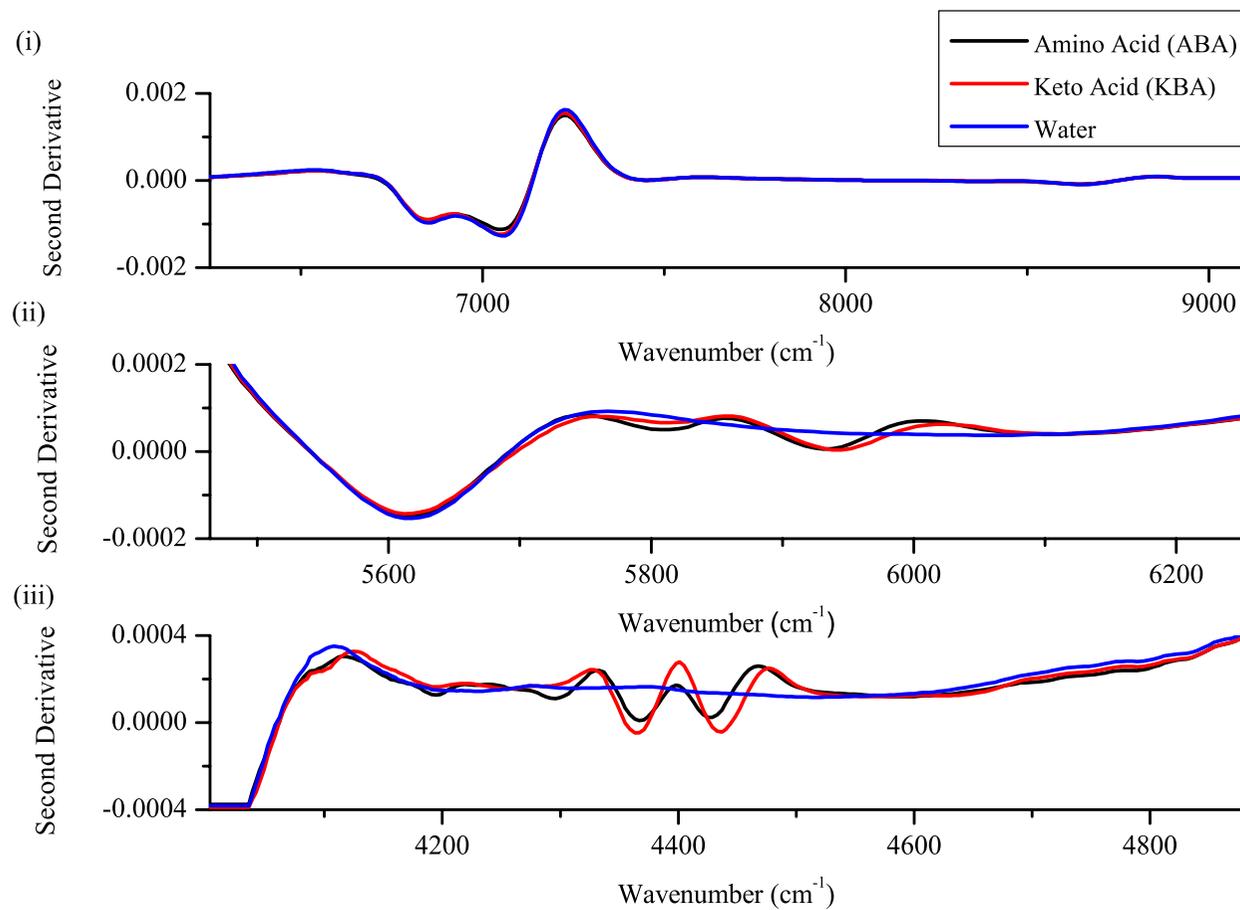


Figure 5-4 - Second derivative spectrum of the pure component spectra of ABA, KBA and water obtained using the at-line near infrared spectrometer system. (i) Displays the second overtones region, (ii) the first overtones and (iii) the combination bands region.

Although the diagnostic region should result in a number of signals that could be attributed to each of the analytes of interest the main focus was on the fingerprint region. Signals in this region should be unique for each of the two compounds. Also if the technique were to be applied *in-situ* then the current technology (diamond ATR crystal) would restrict the utilisable spectra to the fingerprint region because of the wavelengths transmittable by the crystal (diamond absorbs between 1900 cm^{-1} and 2300 cm^{-1}).¹⁴² For these reasons, only the spectral region between approximately 900 cm^{-1} and 1600 cm^{-1} was of interest, however the exact regions utilised for modelling were dependent on the pre-processing methods employed and where the spectra converged to provide an appropriate point at which to cut the spectra.

The mid infrared spectra obtained of the pure component ABA and KBA both illustrated a number of unique spectral features (Figure 5-5). The spectra obtained for these compounds were both very unique, with a number of spectral features that could be uniquely attributed to both the analytes of interest. Whilst it was possible to identify unique features for both analytes in the pure component spectra when a mixture, at concentrations typically expected during the biotransformation process, was considered the amino acid signals began to dominate, particularly at low keto acid concentrations, and it became difficult to identify features arising from the keto acid (Figure 5-5).

The acquired fingerprint region spectra were then converted to the first derivative spectra to try and enhance some of the spectral features, particularly in the mixture samples, but also to reduce baseline drift and noise (Figure 5-6).

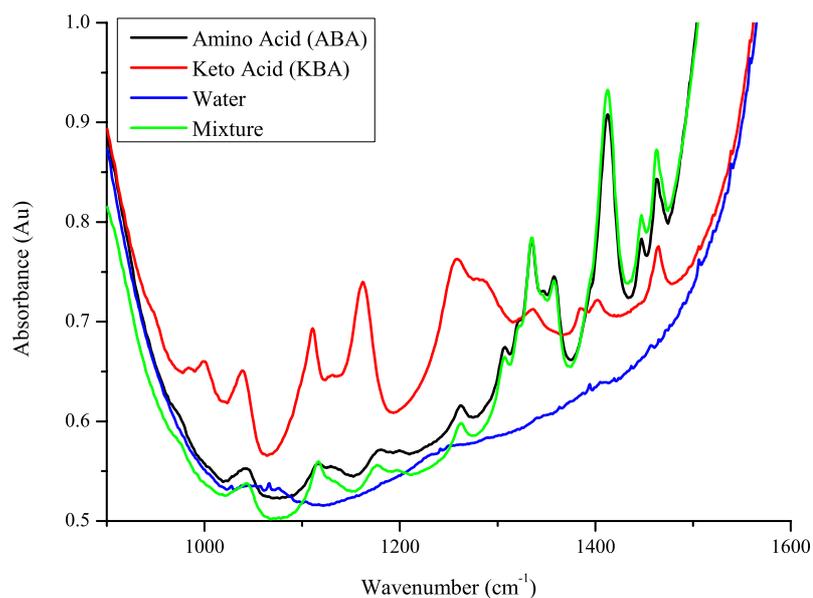


Figure 5-5 - Raw mid infrared spectrum of pure component solutions of amino acid (ABA), keto acid (KBA), water and a mixture of the analytes.

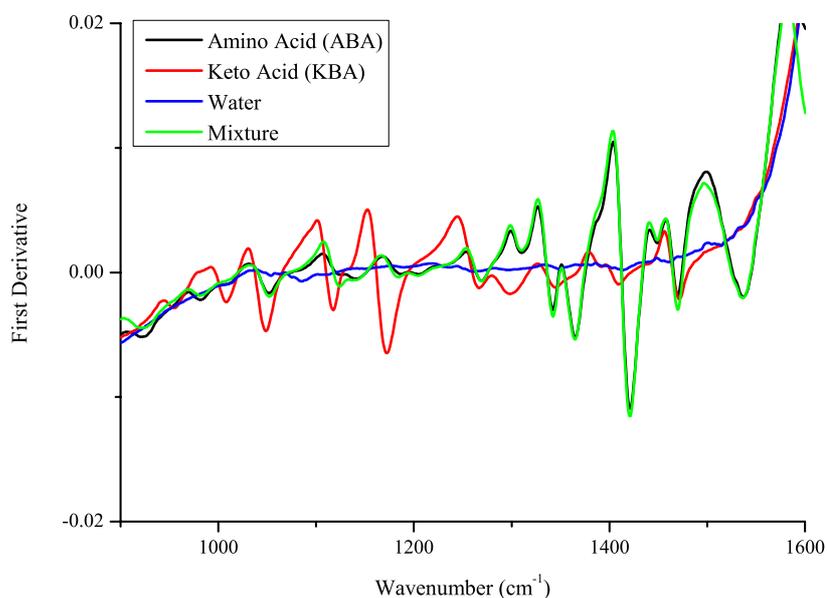


Figure 5-6 - First derivative spectra of pure component samples of amino acid (ABA), keto acid (KBA), the background water matrix and a mixture of the two analytes at concentrations representative of those expected during the biotransformation process.

Looking at the sample mixture of both the amino acid and keto acid it was again quite difficult to identify signals from the keto acid due to the relatively low concentration that was initially present compared with the amino acid, which appears to be dominating the spectrum. Some features observed around the 1100 cm^{-1} & 1300 cm^{-1} areas in the spectrum of the mixture suggested that the keto acid was being identified. It was clear from these spectra of the mixtures that at the concentrations employed a univariate calibration approach would be insufficient to quantify the key analyte concentrations.

5.5 Process Replicates

The feasibility study results suggested that both the near and mid infrared regions were potentially suitable for monitoring this particular biotransformation process. A series of process replicates were thus performed to generate sufficient replicates to construct a calibration model.

One of the key outcomes from the work done with the TBG/TMP biotransformation process replicates was the issue of co-linearity within the resulting model (4.8). With this biotransformation process measures were taken from the outset to ensure that co-linearity within the system was adequately addressed.

The various approaches to breaking co-linearity within the system were evaluated to determine which would be the most appropriate for the given situation. A calibration model based entirely on synthetic samples similar to the approach described by Chung *et al.* (1996) and Riley *et al.* (2001) was considered and dismissed as it was felt the synthetic samples would not adequately represent features of the background matrix such as the cellular material.^{35,37} The semi synthetic approach discussed by Riley *et al.* (1997) and Cervera *et al.* (2009) was also considered as an option, but was again rejected due to the requirement for the samples to undergo an adulteration stage post sampling, therefore rendering the sample inconsistent with the background media present in the biotransformation process.^{17,36}

Previously breaking co-linearity within the system appeared to have been successfully dealt with by varying the analyte concentrations from the start of the biotransformation process, an approach described by Ehyl *et al.* (2007).¹³⁶ This approach had the advantage that the samples were not altered post-sampling and were therefore representative of the biotransformation process.

With the TBG/TMP biotransformation process the batch nature of the substrate preparation created these variations by chance, however in this case, since the substrate was a commercially available pure material, a more structured approach was required.

Since the substrate was introduced as the racemic amino acid, it was necessary to actively ‘spike’ the system with some of the keto acid product, as this would not naturally be present. The starting concentrations of ABA and spiked KBA were determined according to a simple two factor, two level [2²] experimental design (Table 15).

Table 15 - Initial concentrations of amino acid and keto acid determined by experimental design in an attempt to break any co-linearity within the system.

Replicate	Amino Acid (ABA) Concentration (mMol dm⁻³)	Keto Acid (KBA) Concentration (mMol dm⁻³)
1	800	50
2	1200	50
3	800	150
4	1200	150

The four replicates of the biotransformation process were carried out with the initial substrate loadings and spiked keto acid concentrations as determined by the experimental design. Samples were removed at various points during the time course of the process and both the near and mid infrared spectra acquired, in addition the concentrations of both analytes were quantified using the developed HPLC methods (3.2)

As the biotransformation process progressed, changes in the spectral features identified were observed in both the near and mid infrared regions of interest.

A further two replicates of the process were carried out with initial amino acid concentrations in the region of $1000 \text{ mMol dm}^{-3}$ and a KBA spike of approximately 50 mMol dm^{-3} . In terms of concentration, these replicates were at an intermediate ABA concentration and so were within the calibration models predictive ability, but also were of a typical concentration at which the process would normally be run.

5.6 Near Infrared

Partial least squares (PLS) calibration models were constructed for each of the identified regions of interest for both the near and mid infrared spectra. Internal validation of the models was achieved using the leave one out cross validation approach (2.5.5.1). Each of the models was also externally validated using the two unseen process replicates (2.5.5.2).

5.6.1 Calibration Model & Validation

From the pure component spectra acquired of the analytes (Figure 5-4) signals, that were potentially useful for constructing a model for the biotransformation process, were identified in both the first overtones region and the combination bands region. Initially, these regions were considered separately with an individual calibration model constructed for the combination bands region and the first overtones region. A PLS model was then constructed based on the combined spectra from both these regions. External validation of the models constructed was achieved using the two unseen process replicates (biotransformation replicates 5 and 6) and the error values compared (Table 16).

Models based on the combination bands region returned very good errors of calibration and errors of cross validation for the KBA and also reasonable values for ABA. The errors of prediction from the two unseen validation datasets were in good

agreement with the errors of calibration and cross validation for both the amino acid and keto acid. Plots of the measured concentration (from reference analysis methods) plotted against the predicted concentration (from the PLS model) were constructed for both the TBG (Figure 5-7(i)) and the KBA (Figure 5-7(ii)). Points on these plots appeared randomly scattered around the central (1:1) diagonal line suggesting there was not a bias to either over or under predicting the concentrations of the key analytes.

Table 16 – Summary table of the errors associated with the calibration, internal and external validation for the constructed PLS models for the various identified spectral regions of interest.

Region Utilised	RMSEC (mMol dm ⁻³)		RMSECV (mMol dm ⁻³)		RMSEP (mMol dm ⁻³)	
	Amino Acid	Keto Acid	Amino Acid	Keto Acid	Amino Acid [r ²]	Keto Acid [r ²]
Combination Bands (2032-2400 cm ⁻¹)	38.7	5.7	39.1	5.8	(v) 40.7 [0.90] (vi) 35.5 [0.93]	(v) 7.8 [0.99] (vi) 5.6 [0.99]
First Overtones (1540-1812 cm ⁻¹)	40.9	20.3	41.7	20.8	(v) 48.7 [0.85] (vi) 54.6 [0.88]	(v) 14.9 [0.95] (vi) 20.7 [0.90]
First Overtones & Combination Bands (1540-1812 cm ⁻¹ & 2032-2400 cm ⁻¹)	38.1	8.4	38.8	8.6	(v) 41.5 [0.89] (vi) 34.3 [0.93]	(v) 7.0 [0.99] (vi) 6.2 [0.99]

The obtained RMSEP values for the ABA component appeared high in comparison with the values obtained for the KBA component. To put these errors into context they were compared with the errors of prediction values tabulated by Cervera *et al.* (2009) for key analytes of interest in a variety of bioprocess systems. Many of these

errors were quoted in units of g/L therefore it was necessary to convert the obtained error value of 40 mMol dm⁻³ into these units for comparison. An ABA RMSEP value in the region of 40 mMol dm⁻³ equated to 0.0013 g/L.

Key fermentation media constituents such as glucose and glutamine have had reported errors of prediction ranging from 0.072 g/L through to 2.27 g/L using both at-line and *in-situ* spectral acquisition mechanisms.¹⁷ Compared with these values the obtained error for the key constituents of the biotransformation process was very good.

Near infrared spectroscopy utilised to monitor biotransformation processes had previously returned errors of prediction as low as 0.09 g/L with an at-line sampling approach in a Baeyer-Villiger type biotransformation process.⁹³ This error value was higher than had been observed with the de-racemisation biotransformation process under investigation.

Amino acid quantification using near infrared spectroscopy has also been documented in the context of insect cell cultivation. Riley *et al.* (1998) reported errors of prediction as high as 44.2 mMol dm⁻³ for alanine concentrations in this system.³⁸ These errors were comparable with those obtained from the near infrared model of the biotransformation process. This was encouraging as the system currently under investigation represented a more complex and challenging matrix compared with the relatively simple and optically clear insect cell culture systems.²¹

Although the errors of prediction for the amino acid component in particular seem high, relative to the typical errors of prediction quoted for similar systems they were comparable, if not better in many cases.

Models based on the first overtones region of the near infrared spectra returned errors of calibration and cross validation that were comparable with those previously obtained for the amino acid, however the values for the keto acid were substantially higher. Examination of the plot of measured concentration against predicted

concentration illustrated that the predictions of the amino acid concentration were relatively randomly scattered although there was a greater degree of error associated with these samples (Figure 5-7(iii)). The plot for the keto acid concentrations did however suggest that there appeared to be a tendency of the constructed model to over predict KBA concentration relative to the off-line analysis method (Figure 5-7(iv)). This may be a result of the limited variation of spectral features in this particular region. Further investigation into the contributors to the resulting model would be required to determine what spectral features were being modelled and if the inference of KBA concentration from spectral features of the ABA was occurring and hence the sharp increase in the errors of calibration, cross validation and prediction.

Combining both the regions of interest, which also had the effect of removing the dominant signals from water, also resulted in a calibration model that returned errors that were consistent with what had been observed for the individual regions as well as errors of prediction that were in good agreement with the calibration and cross validation errors (Table 16). Plots of measured concentration against predicted concentration once again suggested a random scatter around the central diagonal for both ABA and KBA suggesting there was no bias to either over or under predicting concentrations in the resultant model (Figure 5-7(v) and Figure 5-7(vi)).

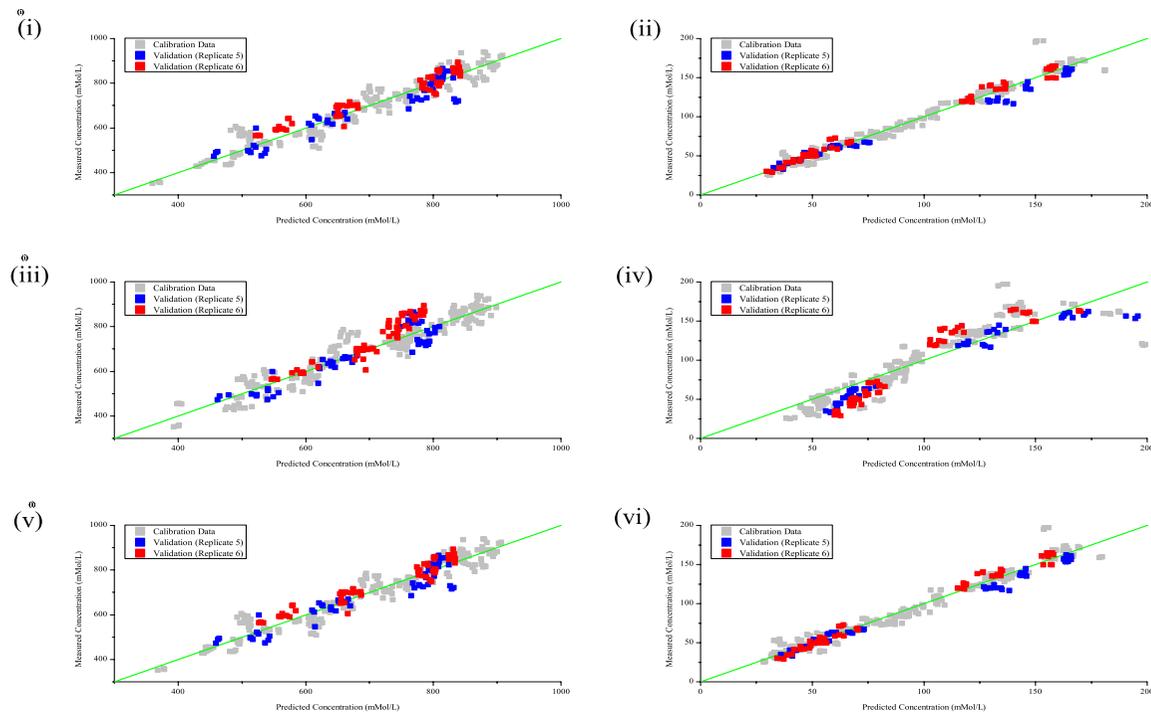


Figure 5-7 – Plots of measured concentration (mMol dm^{-3}) (from reference analysis method) against predicted concentration (mMol dm^{-3}) (from PLS model) for the two validation datasets using the various spectral regions of interest. Figures (i) and (ii) represent the model constructed from the combination bands region, (iii) and (iv) the first overtones region, (v) and (vi) the model constructed from the combined regions. Points shown in grey are the measured concentrations plotted against predicted from the internal validation procedure

5.6.2 Model Contributors

In order to establish what spectral features each of the constructed models was based on, and to ensure that the co-linearity within the system had been effectively dealt with, the regression coefficients for each model were examined.

5.6.2.1 Combination Bands

The two regression coefficients associated with the model based on the combination bands region were plotted alongside the pure component spectra of both ABA and KBA to determine which of these features were contributing to the model (Figure 5-8). As previously noted, the pure component spectra of ABA and KBA were very similar with subtle shifts in peak position observed with a single unique feature assignable to the ABA. These features were expected to be represented and identifiable in the regression coefficients.

The first regression coefficient contained some features that could be assigned to the amino acid (Figure 5-8 (i)). A unique signal for the amino acid observed at 4293cm^{-1} appeared to correspond with the large, negative correlation observed between $4276\text{-}4291\text{ cm}^{-1}$. The subtle variation in peak position between ABA and KBA was reflected in the regression coefficient in the 4454 cm^{-1} region, which in this case appeared to demonstrate a positive correlation with the positive peak of the amino acid. This would therefore suggest that this regression coefficient was representative of the amino acid component.

Strong signals on the second regression coefficient appeared to correspond with the position of second derivative signals in the pure component KBA spectrum (Figure 5-8 (ii)). Troughs in the second derivative pure component KBA spectrum at 4363 cm^{-1} and 4435 cm^{-1} were in good correlation with the negative signals of the regression coefficients plot noted at 4374 cm^{-1} and 4444 cm^{-1} respectively. The negative signal in the loadings plot observed at 4310 cm^{-1} might be representative of the signal unique to the amino acid observed at 4293 cm^{-1} .

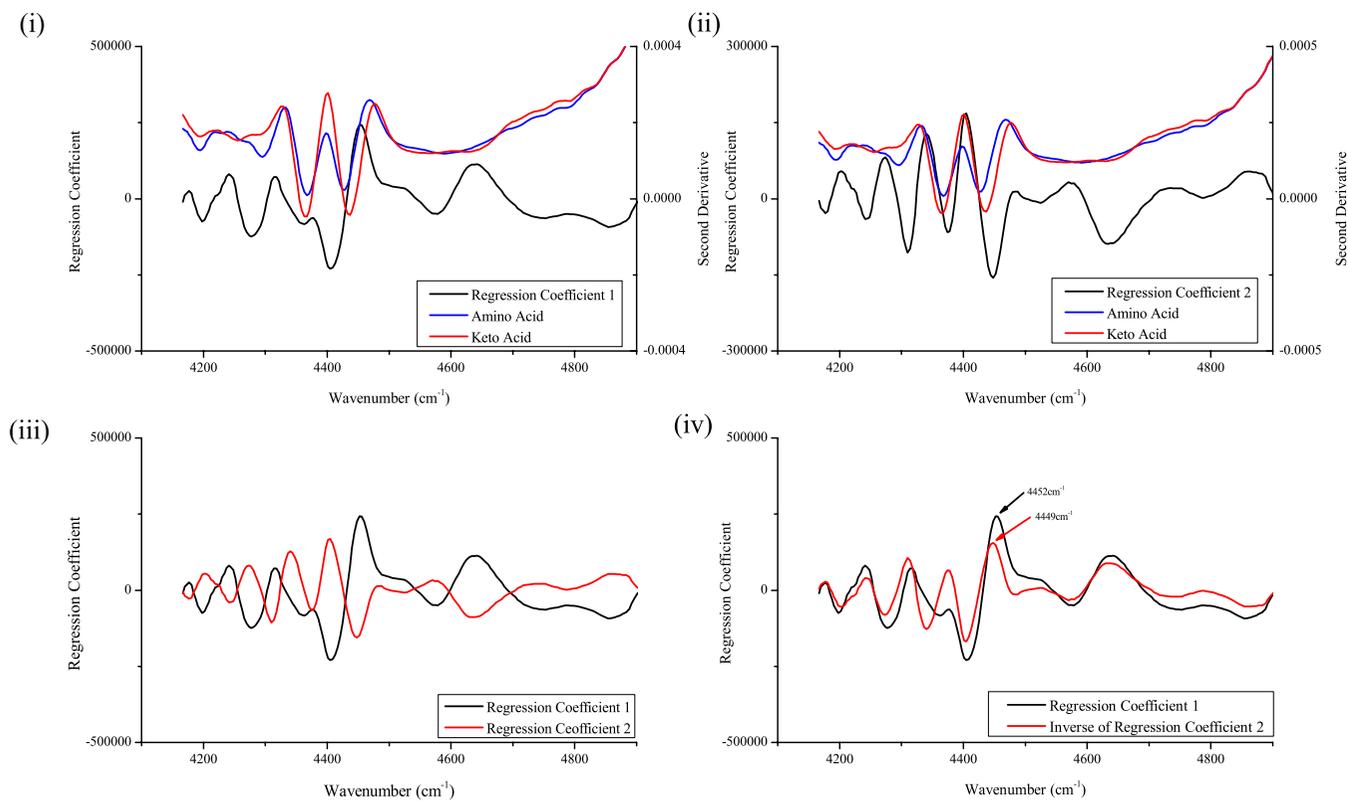


Figure 5-8 - (i) Plot of first regression coefficient and pure component spectra of ABA and KBA. (ii) Plot of second regression coefficient and pure component spectra of the key analytes of interest. (iii) Both regression coefficients obtained for the model plotted for comparison. (iv) Plot of first regression coefficient and inverse of second regression coefficient to highlight subtle variations in the regression coefficients, which result from the subtle shifts in the pure component spectra.

The presence of features attributable to the amino acid in this regression coefficient was a cause for concern as it suggested the co-linearity issue within the system had not been adequately addressed. Despite this potential issue it did appear that the second regression coefficient correlated with the key features of the keto acid spectrum.

If co-linearity within the system were an issue, the regression coefficients would be expected to contain a number of common features. However as both molecules had a very similar structural backbone a number spectral features common to both analytes were expected. Both the regression coefficients obtained for the PLS model were plotted together and examined to identify if any common, overlapping features could be identified (Figure 5-8 (iii)). Interestingly, the regression coefficients appeared to be a mirror image of each other, with the unique signal for the amine functional group appearing to have an influence in both regression coefficients.

As previously noted, the shifts in the peak positions of the two analytes were very subtle. This was expected to be reflected in the regression coefficients, and may make them appear, visually at least, mirror images of each other. In order to assess how similar or otherwise the regression coefficients were the inverse of the second regression coefficient was plotted alongside the first regression coefficient (Figure 5-8 (iv)). This essentially rotated the second regression coefficient through 180° so that the plots could be superimposed for comparison purposes.

The plot of the first regression coefficient and the inverse of the second regression coefficient illustrated that there were differences between the two regression coefficients obtained for the model. Shifts in the peak positions were noted over various regions; particularly noteworthy were the signals at 4449 cm⁻¹ and 4454 cm⁻¹. This was one of regions identified as exhibiting subtle variance between the two key analytes.

Spectra used to identify various features in the regression coefficients were pure component samples, presented at median concentrations (~500 mMol dm⁻³ ABA and

~250 mMol dm⁻³ KBA) expected during the biotransformation process. Samples used for the construction of the model contained a mixture of the analytes as well as cellular material and other, undefined matrix components. Some features of the regression coefficients may be representative of these components resulting in additional, unassigned signals and shifts due to these more complex matrices.

Although both regression coefficients were similar, due to the structural similarities of the molecules and the resulting infrared spectra, the subtle variations noted in the pure component spectra were reflected in the regression coefficients. Two different regression coefficients were obtained suggesting co-linearity was not an issue, however ideally some representation of the unique amine feature in only the first regression co-efficient would have confidently confirmed the models ability to independently quantify each analyte.

5.6.2.2 First Overtones

The PLS model constructed based on the first overtones region of the near infrared spectra did not return errors of calibration, cross validation or prediction that were comparable with those previously obtained based on the combination bands region. Given the limited spectral region and the close similarities observed between the spectra of the two analytes of interest in this region these higher errors were not unsurprising. This close similarity between the spectra of the two analytes also made interpretation of the regression coefficients associated with the model difficult since the differences were based on a small shift in peak position.

The first regression coefficient associated with the model was plotted alongside the pure component ABA and KBA spectra (Figure 5-9 (i)). A number of features of this regression coefficient were aligned with signals in the pure component spectrum of the ABA, signals at 5800 cm⁻¹ and the position of the signal at 5850 cm⁻¹ correspond well with strong negative values in the regression coefficient.

In this case, the second regression coefficient obtained for the model appeared very different from the first (Figure 5-9 (ii)). Although a very weak signal was observed in the pure component spectra of KBA at 5800 cm^{-1} , there did not appear to be any corresponding signal in the regression coefficient. The main feature of this regression coefficient was observed just below 6000 cm^{-1} and represented the KBA signal that exhibits an overlap with an ABA signal in the same region but with its peak apex at a slightly lower wavenumber.

Both the regression coefficients were plotted together for comparison (Figure 5-9 (iii)). In this case, a number of unique features were observed for each regression coefficient as well as some features that appeared to overlap, which were expected due to the similarities of the analytes.

The inverse of the second regression coefficient was determined and plotted alongside the first regression coefficient (Figure 5-9 (iv)). This identified if there were any subtle differences observed in the signal positions, or if there was an overlap in a correlated/anti-correlated manner. It was clear from this figure that there was an obvious difference in the position of any overlapping signals. Of particular note was the large positive value in the regression coefficient at 5980 cm^{-1} , that appeared to correspond with a positive signal in the second derivative spectrum of the amino acid in this region. This signal overlapped with a large negative regression coefficient at 5976 cm^{-1} however this appeared to correspond, slightly shifted, with the second derivative trough for the keto acid observed at approximately 5900 cm^{-1} .

Since two distinct regression coefficients were observed, it was concluded that the co-linearity within the system had been suitably addressed to allow the constructed model to independently predict the concentration of both ABA and KBA.

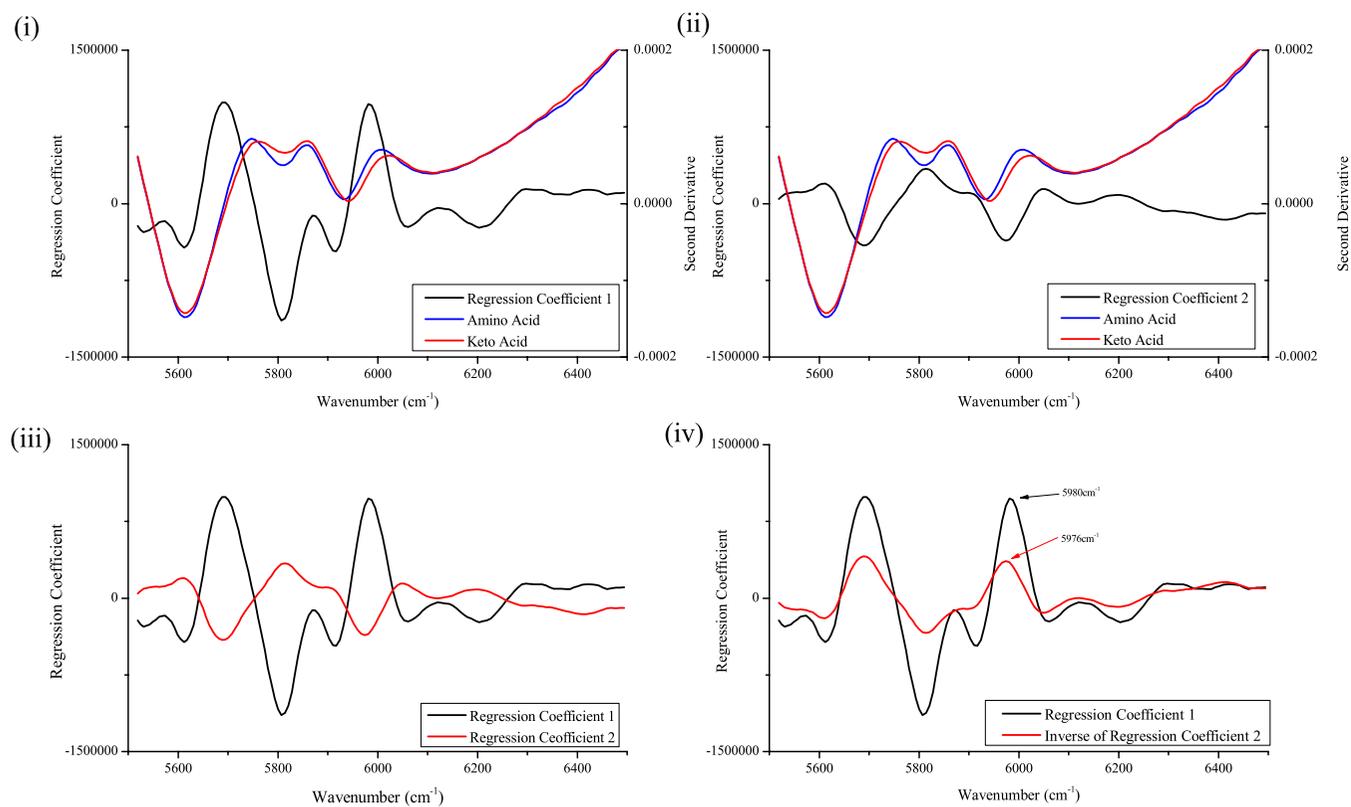


Figure 5-9 – (i) Plot of first regression coefficient and pure component spectra in first overtones region. (ii) Second regression coefficient along with pure component spectra in the first overtones region. (iii) Both regression coefficients associated with the model plotted together for comparison. (iv) Plot of first regression coefficient and inverse of second regression coefficient to identify any subtle shifts in the regression coefficients.

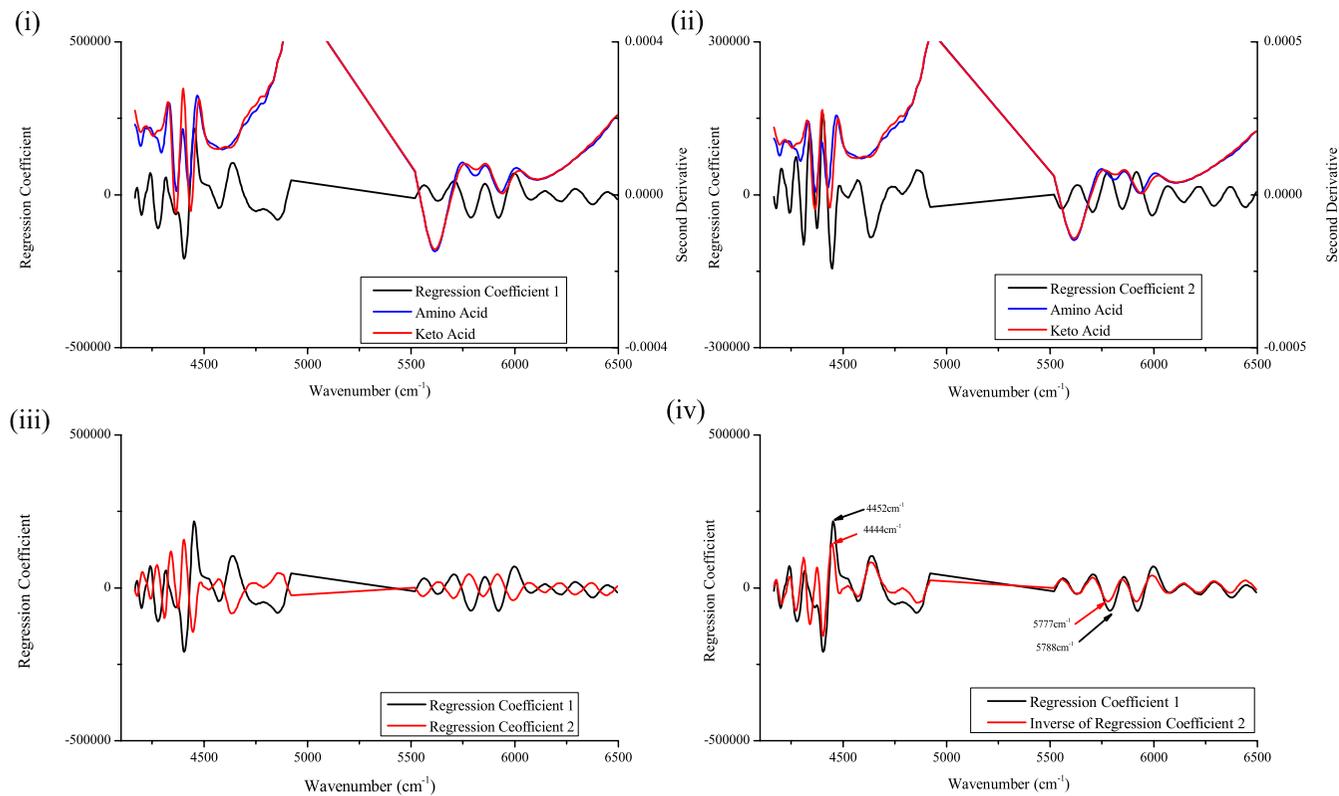


Figure 5-10 - (i) Plot of first regression coefficient along with second derivative spectrum of pure component ABA and KBA. (ii) Plot of second regression coefficient and pure component ABA and KBA samples. (iii) Regression coefficients associated with the constructed model plotted for comparison. (iv) First regression coefficient and inverse of second regression coefficient overlaid for comparison of peak positions.

5.6.2.3 Combined Model

Combining both spectral regions of interest generated a third PLS model for this biotransformation process. Errors of calibration, cross validation and prediction observed for this model were some of the best noted for the amino acid (ABA) but also provided a reasonable quantification of the keto acid (KBA) content (Table 16).

As with the models based on only the first overtones and combination bands regions, the regression coefficients associated with the model were examined to try and confirm whether or not the co-linearity within the system had been addressed. Since this model was based on the combined regions many of the spectral features previously identified in each regression coefficient were expected to be duplicated in this system.

The first regression coefficient contained features that were consistent with those previously described for the individual models for this region (5.6.2.1 & 5.6.2.2). Points that exhibited a strong negative regression coefficient appeared to correspond with the second derivative troughs noted in the pure component ABA spectrum (Figure 5-10 (i)). Of particular note was the unique signal of the amine functional group observed at 4293 cm^{-1} , which corresponded exactly with a large negative point on the regression coefficient.

Features of the second regression coefficient appeared to be more aligned with those of the pure component keto acid, particularly around the 4000 cm^{-1} to 4500 cm^{-1} region where features of the pure component spectra essentially overlay with the second regression coefficient (Figure 5-10 (ii)). However the regression coefficients in the first overtones region were different from those observed in the model based on this sole region (5.6.2.2). There also appeared to be some influence of the amine functional group of the ABA in this second regression coefficient.

When both regression coefficients were plotted together (Figure 5-10 (iii)) a mirror image effect was once again observed. With the combined model however this had

extended to include the first overtones region, which had previously returned two very distinct regression coefficients.

Comparison of the first regression coefficient with the inverse of the second regression coefficient confirmed there was substantial overlap (Figure 5-10 (iv)). Whilst there were some subtle variations in the position of the peaks, the vast majority of the regression coefficients were identical. This would suggest this model was not independently modelling both analytes of interest.

5.6.2.4 Near Infrared Summary

Models based on the near infrared spectra had been generated for the process, and were relatively successful at predicting the analyte concentration in unknown samples. Examination of the contributors to these models appeared to suggest there was some degree of independence surrounding the models ability to predict analyte concentrations. In this case, the most successful model incorporated the combination bands region of the near infrared spectrum. Hence, the technique would be restricted to a rapid at-line analysis method. Whilst this was a substantial improvement over the current off-line chromatographic methods, an *in-situ* approach would still be the preferred option.

Additionally, the close similarities between the spectra of the two analytes of interest still raised questions over the ability of the constructed model to independently quantify these two key analyte concentrations. The regression coefficients for the various models suggested that the spectral features used to estimate the concentration of each analyte were similar. Of particular note was the combined model where the regression coefficients were exact opposites of each other. Results like this would suggest that co-linearity had not been adequately addressed or that there were insufficient spectral features to allow construction of a robust model.

The mid infrared region was therefore considered as an option to monitor this particular biotransformation process. As with the de-racemisation of TBG/TMP

biotransformation process the inclusion of the fingerprint region of the mid infrared by definition should result in a greater variation in the acquired spectra to allow construction of a model with greater confidence in its ability to independently predict the analyte concentrations.

5.7 Mid Infrared Spectroscopy

5.7.1 Mid Infrared Raw Spectra

Scans in the mid infrared region of the pure component samples of ABA and KBA (Figure 5-5) suggested that it may have been possible to utilise the raw mid infrared spectra to construct a quantitative model for the biotransformation process.

The initial feasibility study suggested that the spectral region between 900 cm^{-1} and 1600 cm^{-1} would be a suitable region for use in the modelling process. Further evaluation of both the raw spectral data and the first derivative data suggested that suitable points to ‘cut’ the acquired spectra were at 940 cm^{-1} and 1465 cm^{-1} retaining the intermediate spectral range.

The spectra from the four ‘spiked’ replicates of the biotransformation process designated as calibration datasets were utilised to construct a PLS model for the process. In the first instance the only pre-processing applied to both the spectral and reference data blocks was to mean centre the data. Internal validation using the LOOCV approach (2.5.5.1) was also applied to the constructed model.

The optimum number of latent variables utilised in the model was determined from the plots of RMSEC and RMSECV values for both the analytes of interest. From this plot (Figure 5-11) the optimum number of latent variables appeared to be five. This constructed model returned error values as detailed in Table 17.

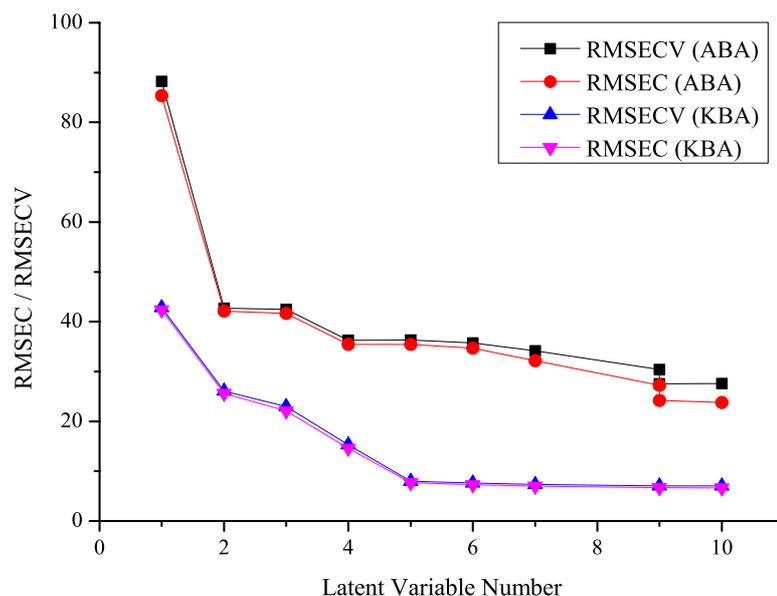


Figure 5-11 - Plot of RMSEC / RMSECV values against latent variable number used to determine the optimum number of latent variables for the constructed PLS model based on the mid infrared spectra without any spectral pre-processing applied.

Table 17 - Summary of error values for the calibration, internal and external validation procedures for the PLS model constructed from the raw mid infrared fingerprint spectra.

	ABA (mMol dm ⁻³) [r ²]	KBA (mMol dm ⁻³) [r ²]
RMSEC	35.4	7.7
RMSECV	36.3	8.0
RMSEP	(i) 38.7 [0.90] (ii) 39.5 [0.92]	(i) 9.1 [0.99] (ii) 15.2 [0.98]

The errors from the internal validation procedure were in good agreement with the errors of calibration for the model. Although encouraging, this was not unsurprising given that the internal validation procedure uses data to test the model that has essentially been used in the calibration procedure.

To challenge the model spectra, from both the biotransformation replicates were presented to the constructed model and the concentrations of the two key analytes in these samples predicted based on the spectral data. These values were compared with the measured values determined using the off-line chromatographic techniques (3.2) to assess how well the model was performing.

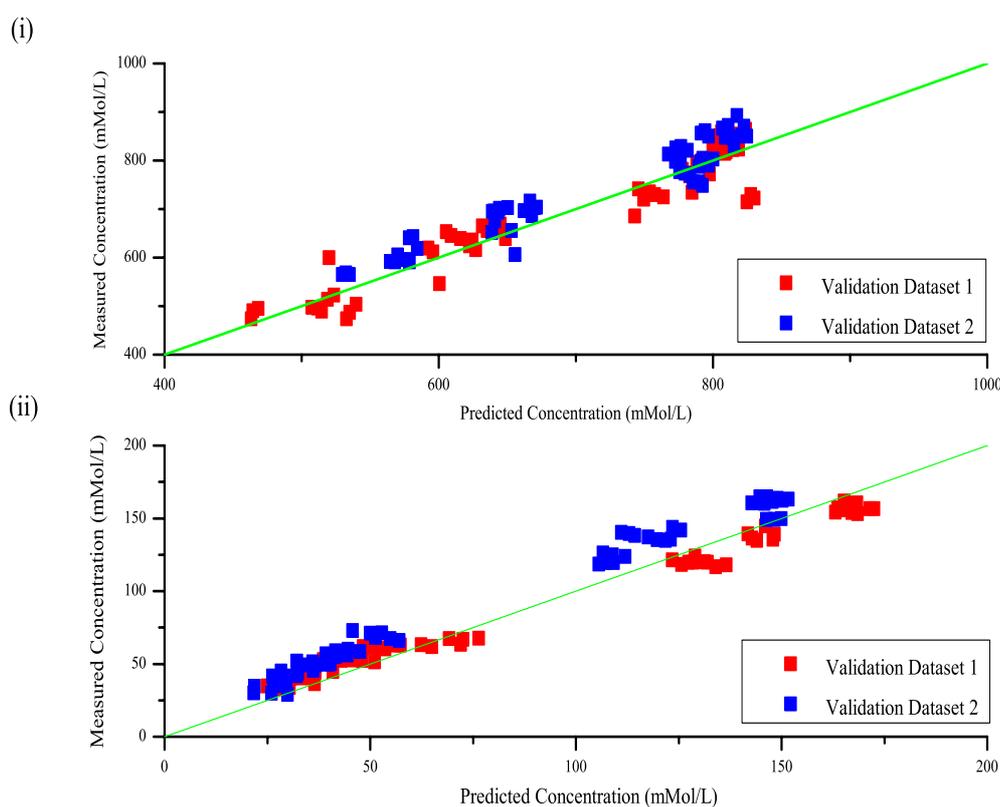


Figure 5-12 - Plots of measured concentration (mMol dm^{-3}) (from reference analysis method) against predicted concentration (mMol dm^{-3}) (from the PLS model) for both the amino acid (i) and keto acid (ii) obtained from the two external validation datasets.

The errors of prediction for the amino acid component were in good agreement with the errors of calibration and cross validation. Evaluation of the scatter plots of measured concentration against predicted concentration (Figure 5-12) suggested that the predicted concentrations were in reasonable agreement with the measured

concentrations for the amino acid. Points appeared to be randomly scattered around the central diagonal line representing a 1:1 relationship suggesting the model wasn't bias towards over or under predicting the analyte concentration. With respect to the KBA concentrations, the first validation dataset appeared to predict the concentrations very well, with the points randomly scattered around the central diagonal. The second validation dataset however returned a much higher error of prediction and on examination of the plot of measured concentration against predicted concentration there appeared to be a very slight tendency of the model to under predict the keto acid concentration relative to that determined by the analytical reference method.

In order to identify the contributors to the constructed model, and to confirm whether the suspected co-linearity within the system had been addressed, the regression coefficients associated with the model were examined.

The unmodified pure component mid infrared spectra of ABA and KBA were plotted alongside the first regression coefficient (Figure 5-13 (i)). A number of points in the regression coefficient were matched with signals in the pure component spectrum of ABA. Signals observed at 1400 cm^{-1} and 1350 cm^{-1} corresponded with large values of the same sign on the regression coefficient. In addition the pattern of signals noted between 1300 cm^{-1} and 1350 cm^{-1} corresponded with a similar pattern in the regression coefficient with the opposite sign. These observations suggested that the first regression coefficient was related to the ABA component. A large negative regression coefficient observed just above 1100 cm^{-1} appeared to inversely correspond with one of the signals observed in the pure component KBA sample.

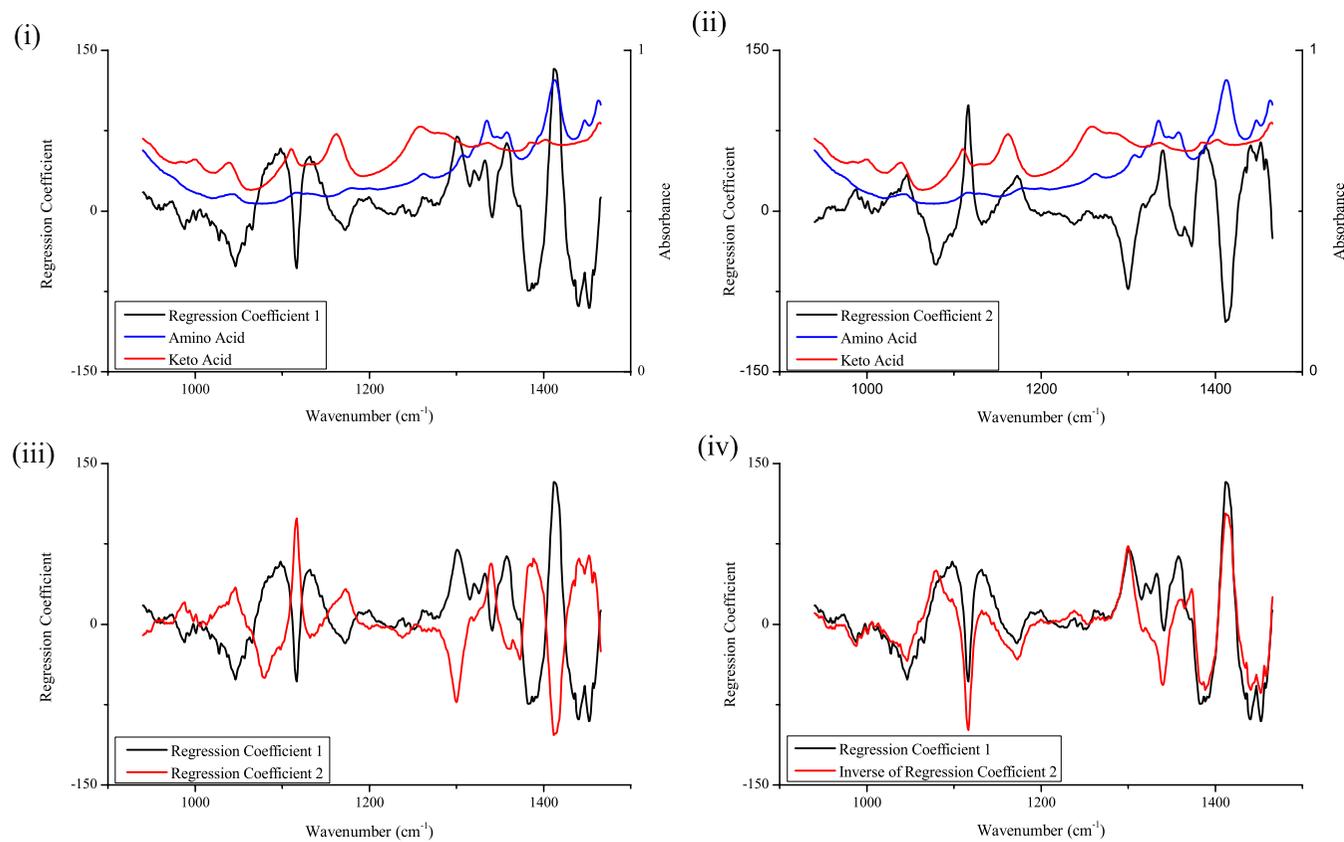


Figure 5-13 - (i) Regression coefficient 1 plotted alongside the pure component spectrum of ABA and KBA. (ii) Second regression coefficient plotted with pure component ABA and KBA spectra. (iii) Both regression coefficients associated with the constructed model superimposed. (iv) First regression coefficient and inverse of second regression coefficient superimposed.

The presence of this particular signal in the first regression coefficient was a concern as it suggested some influence of the spectral features due to KBA being utilised in the estimation of ABA concentration.

A plot of the second regression coefficient and the pure component ABA and KBA samples was also constructed and examined (Figure 5-13 (ii)). Positive values for the regression coefficient in the 1100 cm^{-1} to 1200 cm^{-1} region corresponded with the signals observed in the pure component KBA spectrum. However the strong signal at 1400 cm^{-1} that represented ABA was also noted in this regression coefficient in an anti-correlated manner. The presence of these features in the regression coefficient suggested there could still be some co-linearity issues within the model.

Both the regression coefficients were plotted together for comparison (Figure 5-13 (iii)). As had been the case with many of the near infrared models, the regression coefficients appeared to be mirror images of each other. The inverse of the second regression coefficient was plotted alongside the first regression coefficient to try and identify any subtle variations that may be observed (Figure 5-13 (iv)). Both regression coefficients were very similar, some differences were noted in the 1300 cm^{-1} to 1400 cm^{-1} , but other than these the regression coefficients were identical.

These observations with the regression coefficients brought into question whether the model was able to independently quantify the concentration of both ABA and KBA.

5.7.2 Mid Infrared First Derivative

Although a model for the biotransformation process had been constructed based on the spectral data without any mathematical manipulation, the errors of prediction, for the keto acid in particular, were higher than expected and it was unclear whether the issue of co-linearity had been adequately addressed.

The acquired spectra were converted to the first derivative format and the models re-constructed in an attempt to improve the errors associated with the model.

As had been the case previously, the four replicates of the biotransformation process that had the initial ABA and KBA concentrations ‘spiked’ according to an experimental design were used as the calibration datasets. The optimum number of latent variables for the model was selected as four based on the plots of RMSEC and RMSECV value against the number of latent variables in the model.

External validation of the constructed model was achieved using the two unseen process replicates in the same manner as before with the error values tabulated.

Table 18 - Error values associated with the PLS model constructed from the first derivative fingerprint region mid infrared spectra.

	ABA (mMol dm⁻³) [r²]	KBA (mMol dm⁻³) [r²]
RMSEC	37.4	7.8
RMSECV	38.3	8.2
RMSEP	(i) 41.5 [0.90] (ii) 30.1 [0.92]	(i) 13.0 [0.99] (ii) 9.0 [0.99]

In this case, the errors of calibration and cross validation obtained were comparable with those obtained from the raw spectral data. The errors of prediction, however, were better than had been observed with the raw spectral data. Evaluation of the plots of measured concentration against predicted concentration suggested that the amino acid concentrations were in good agreement with the measured concentrations, randomly scattered around the central diagonal and the concentrations were within the calibration range of the constructed model. The keto acid was acceptable for the first validation dataset however the second dataset appeared again to be slightly under predicting the concentration of KBA (Figure 5-14).

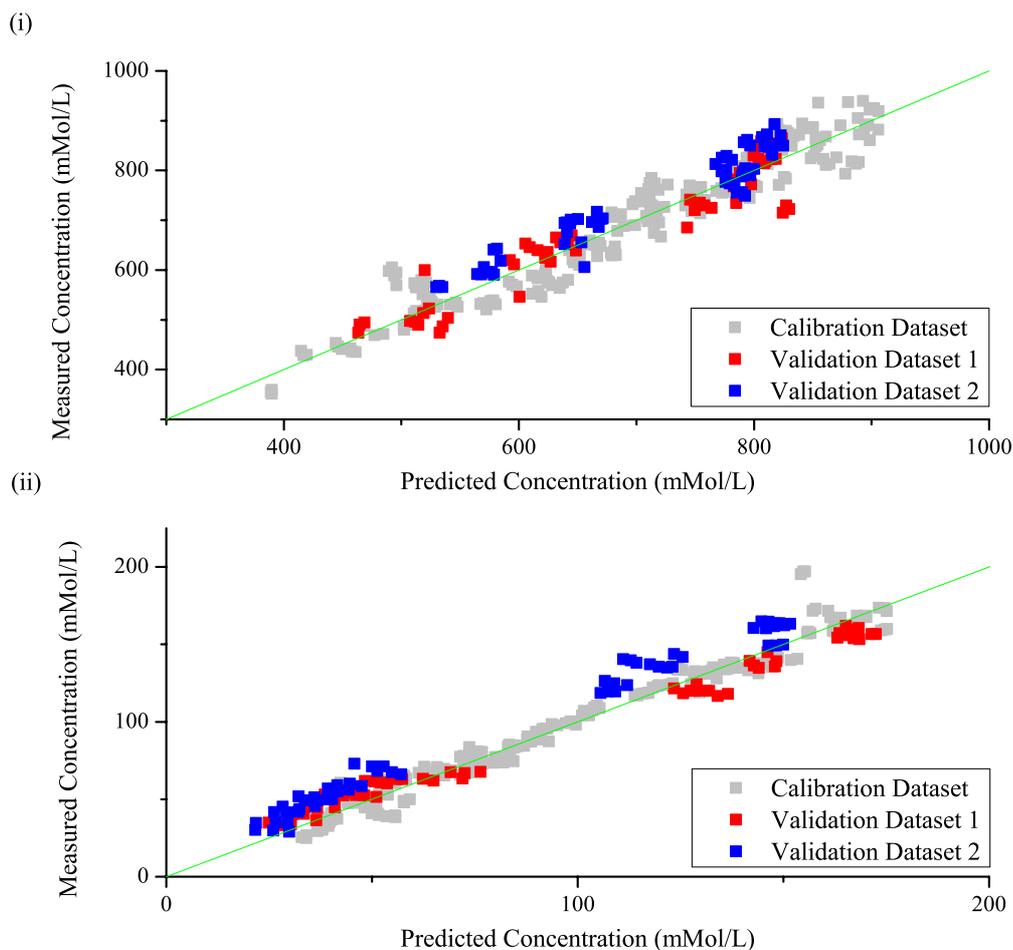


Figure 5-14 – Measured concentration (mMol dm⁻³) (from reference analysis method) plotted against predicted concentration (mMol dm⁻³) (from PLS model) for external validation datasets. (i) Illustrates the amino acid (ABA) component whilst (ii) illustrates the keto acid (KBA) component. Also shown is the measured vs. predicted concentrations from the LOOCV for the calibration data.

With a large difference in the errors of prediction for the two validation datasets between the PLS model constructed from the raw spectra and the first derivative spectra suggested there may be some subtle features in the spectra that are influencing the calibration and analyte concentration predictions.

Contributors to the model were subsequently examined to confirm the intended analytes were being modelled independently. The first regression coefficient associated with the model was plotted alongside the first derivative pure component spectra of ABA and KBA (Figure 5-15 (i)). This first regression coefficient had a number of features that corresponded with the pure component ABA, particularly in the region from 1250 cm^{-1} through to 1400 cm^{-1} where the spectra and regression coefficient were an almost identical match.

With the second regression coefficient signals were observed that corresponded with the pure component KBA (Figure 5-15 (ii)). Based on the pure component spectra the majority of KBA signals were expected in the approximate region between 1150 cm^{-1} and 1250 cm^{-1} . Whilst the second regression coefficient did contain corresponding features in this region, a number of large regression coefficient values were observed in the region from 1300 cm^{-1} through to 1500 cm^{-1} . This region previously appeared to be dominated by signals from the amino acid, however there did not appear to be a good correlation, or anti-correlation, between this second regression coefficient and the pure component amino acid signals. Upon closer inspection, it appeared that calculation of the first derivative had the desired effect and enhanced some of the weaker spectral features of the keto acid within this region. The second regression coefficient appeared to be influenced by some of these weak, overlapping spectral features attributable to the KBA.

Both regression coefficients associated with this model were plotted together for comparison purposes (Figure 5-15 (iii)). From this plot it was clear there were a number of unique features observed in both the regression coefficients. Particular differences were noted in this region above 1300 cm^{-1} , an area that had previously been considered to be dominated by signals from the amino acid. As a confirmation, the inverse of the second regression coefficient was plotted alongside the first regression coefficient (Figure 5-15 (iv)). This clearly illustrated differences, in the form of subtle shifts and unique features, across the full spectral range but in particular the region above 1300 cm^{-1} .

These results not only suggested that co-linearity within the system had been addressed but that by calculating the first derivative of the absorbance spectra some spectral features had been enhanced. These subtle spectral features were having a positive impact on the model, dealing with co-linearity but also improving the errors of prediction with the unseen validation datasets compared with models built using the absorbance spectra.

Observing two independent regression coefficients for the model increased confidence that the model was independently quantifying both ABA and KBA during the biotransformation process. Having determined the optimum spectral pre-processing approach further improvements to the model may result by refinement of the spectral regions used in the model.

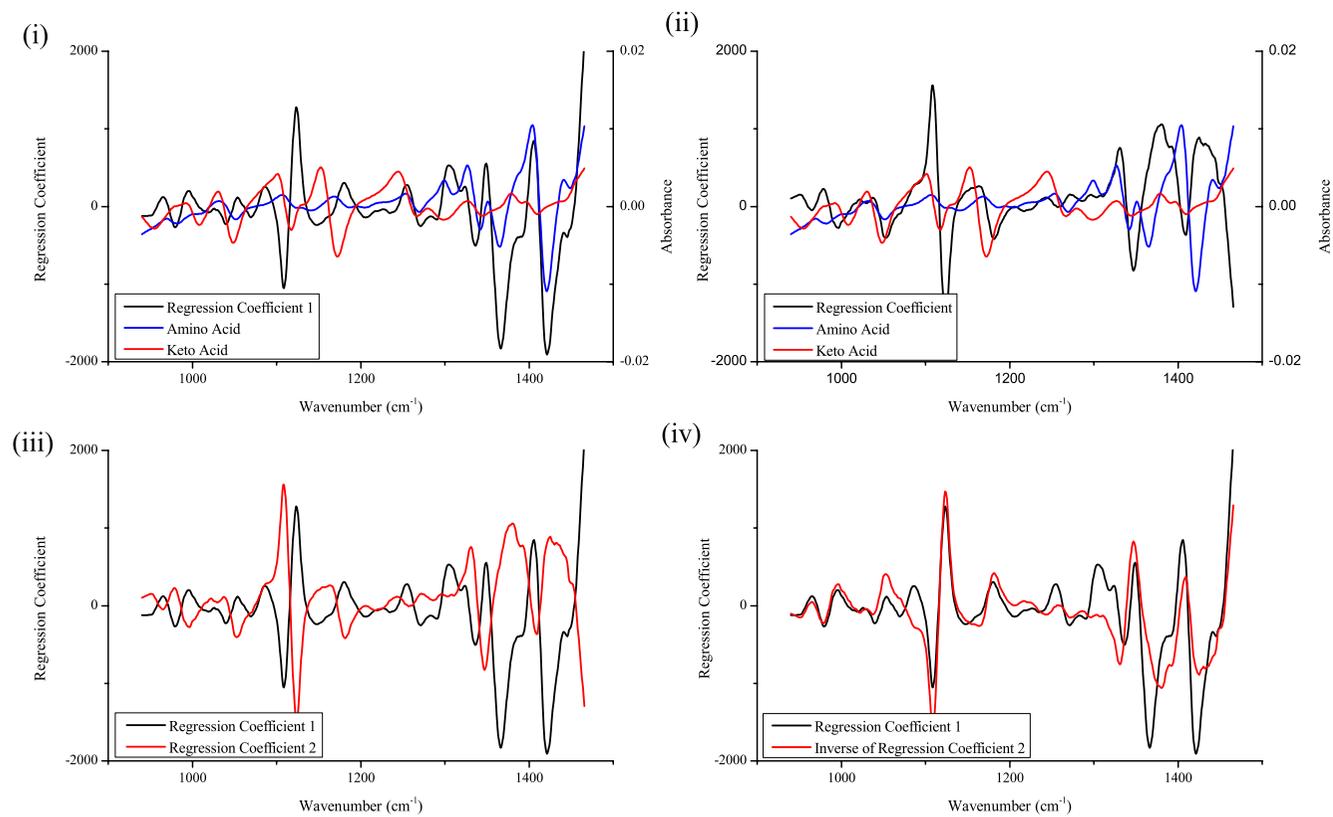


Figure 5-15 - (i) Regression coefficient 1 plotted alongside the pure component spectrum of ABA and KBA. (ii) Second regression coefficient plotted with pure component ABA and KBA spectra. (iii) Both regression coefficients associated with the constructed model superimposed. (iv) First regression coefficient and inverse of second regression coefficient superimposed.

5.7.3 Genetic Algorithm

Spectral regions of interest had been identified visually based on the known location of signals from scans of the pure component material. Further refinement of the spectral region included in the model building process was achieved by utilising a genetic algorithm for variable selection.

All the spectra from the four replicates of the biotransformation process identified as the calibration dataset, and the corresponding reference data, were used with the genetic algorithm to determine the most appropriate spectral regions for monitoring.

Parameters for the genetic algorithm were set for the maximum population size and a window width of 1 so each point in the spectrum (wavenumber) was treated as an independent data point.

5.7.3.1 Raw Dataset

The genetic algorithm carried out on the raw spectra for the calibration dataset identified a total of 139 variables (wavenumbers) from a possible total of 546 variables that should be utilised in the construction of a PLS model. A plot of these selected variables against wavenumber suggested that the variables identified were in the regions where the key signals from both the ABA the KBA were observed (Figure 5-16).

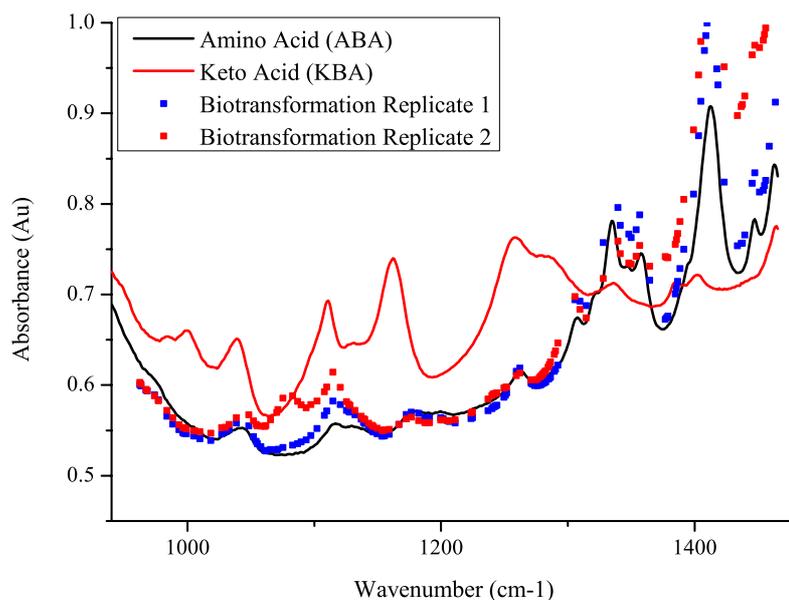


Figure 5-16 - Plot of absorbance measured against variable number for the variables identified by the genetic algorithm for inclusion in the PLS model.

Using only those spectral features identified by the genetic algorithm, the PLS model for the biotransformation process was reconstructed. As had been the case previously (5.7.1) the model utilised the raw spectral data (in absorbance format) with mean centering the only pre-processing technique applied to both the spectral and reference data. In this case, a total of five latent variables were determined as being the most suitable for the model with the errors of calibration and cross validation as quoted in Table 19.

Table 19 - Error values for PLS model constructed using the genetic algorithm.

	ABA (mMol dm⁻³) [r²]	KBA (mMol dm⁻³) [r²]
RMSEC	35.3	8.7
RMSECV	36.2	9.0
RMSEP	(i) 38.6 [0.90] (ii) 37.1 [0.92]	(i) 8.6 [0.99] (ii) 19.4 [0.98]

This attempt to refine the spectral regions and improve the errors of calibration and prediction of the model has failed to deliver the expected improvements. Considering the amino acid (ABA) analyte, the errors of calibration were comparable with those obtained for the full mid infrared raw spectral region (Table 17) with a marginal improvement noted with the errors associated with the predicted concentrations for the two unseen validation replicates. With regards to the keto acid (KBA) concentrations however, the errors of calibration were marginally worse using the variables identified by the genetic algorithm and the errors of prediction were substantially higher than had been observed previously. This was most likely a result of the genetic algorithm not identifying a sufficient number of the KBA's weak spectral features for inclusion in the model.

5.7.3.2 First Derivative Dataset

The genetic algorithm for variable selection was also carried out on the first derivative spectra. Again the intention of this was to try and reduce the spectral features to only those that were essential to try and improve the errors of calibration and prediction associated with the constructed model.

The four process replicates identified as the calibration datasets and the corresponding reference analysis were loaded to perform the genetic algorithm. Algorithm parameters were the same as those that had been previously utilised with the un-derivatised spectra. The genetic algorithm process identified 110 variables from the 546 in the full spectral region for inclusion in the calibration model.

Using this reduced spectral region, the PLS model for the process was re-constructed and the external validation procedure applied to test this constructed model with previously unseen datasets (Table 20). The model was based on three latent variables as determined from the plot of RMSEC/RMSECV against the number of latent variables.

Table 20 - Errors of calibration and prediction associated with first derivative spectra based on regions determined by the genetic algorithm.

	ABA (mMol dm⁻³) [r²]	KBA (mMol dm⁻³) [r²]
RMSEC	33.9	9.3
RMSECV	34.5	9.5
RMSEP	(i) 39.6 [0.90] (ii) 38.2 [0.92]	(i) 9.5 [0.99] (ii) 9.7 [0.99]

This model returned errors of calibration and cross validation that were consistent with those previously determined for both the amino acid and keto acid. Errors of prediction for the two unseen replicates of the biotransformation process were better than had previously been obtained, with both validation replicates returning errors of prediction comparable with the errors of calibration and cross validation.

Examination of the plots of measured concentration against predicted concentration (Figure 5-17) once again suggested that for the amino acid there was random scatter of the points suggesting no bias towards either over-prediction or under-prediction. In this case the keto acid also appeared to be more randomly distributed than had previously been noted. Arguably, there may be a slight tendency towards over predicting the KBA concentration of the second validation dataset. In this case the predicted concentrations lie within the range observed with the calibration data, whereas with previous models points on this dataset have largely been outwith the span of the calibration dataset.

Variables identified using the genetic algorithm of the first derivative middle infrared spectra returned the best errors of calibration, cross validation and prediction for the external datasets. This dataset represented the best model constructed for the ABA/KBA biotransformation process.

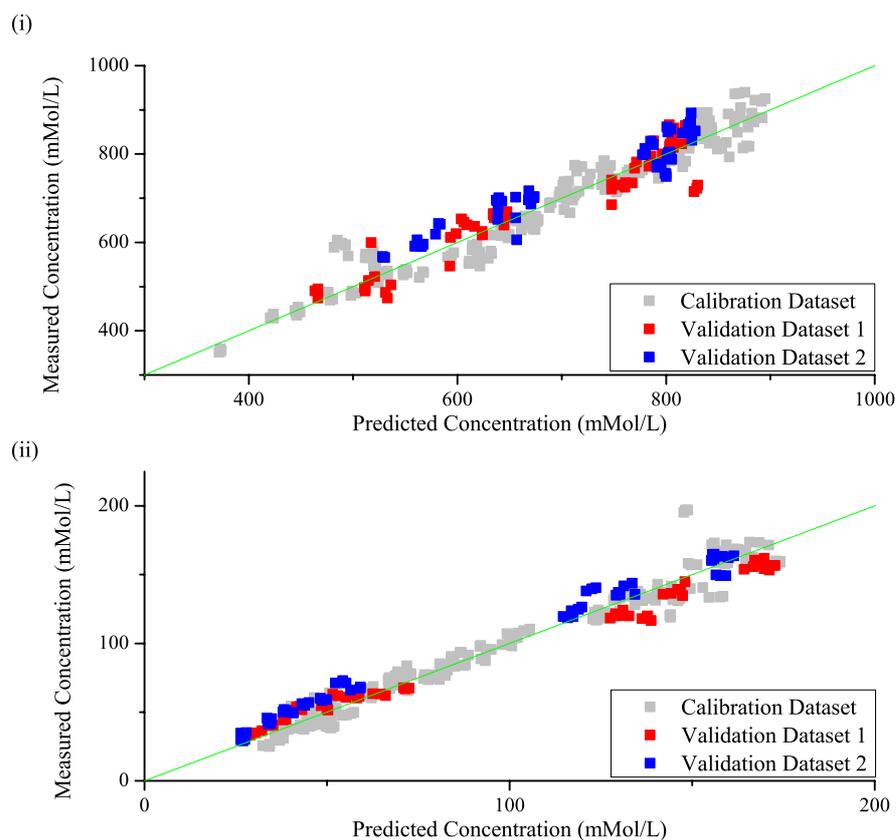


Figure 5-17 – Plots of measured concentration against predicted concentration for two validation datasets for the model built based on the first derivative regions identified by the genetic algorithm.

5.8 Model Summary

A variety of different models have been constructed for the biotransformation process involving the de-racemisation of ABA by conversion of the D enantiomer to KBA (summarised in Table 21). These models have utilised spectral data from both the near and mid infrared regions but also different spectral ranges within these regions have been investigated.

Models based on the near infrared data were constructed and tested for the biotransformation process. However, these models relied heavily on the inclusion of

the combination bands region to construct a model that returned reasonable errors of calibration, cross validation and prediction. The need for this particular region limited the technique to a rapid at-line method, which although an advance over the currently employed method was not the preferred scenario for this particular system.

Models based on the fingerprint region spectra in the mid infrared region were also constructed and tested using the external validation procedure. Various models were built utilising different regions of the fingerprint spectra as well as applying various forms of mathematical manipulation of the spectra. Generally these models all returned consistent errors of calibration and cross validation, that were also in good agreement with the values obtained from the near infrared model.

The best model, that returned the lowest errors of prediction for both of the unseen validation datasets, was that constructed from the first derivative dataset using the variables identified by the genetic algorithm. Although this model had the highest errors of calibration and cross validation of all the mid infrared models constructed, the errors of prediction were better with both datasets returning errors of prediction that were in line with the errors of calibration and cross validation.

Other models constructed from the same data but based on different spectral regions and pre-processing methods were successful at predicting the ABA concentrations of the validation dataset samples, however the predictions of KBA concentrations were not as accurate.

Table 21 – Summary of the models constructed for the quantification of ABA and KBA during the biotransformation processes.

	Spectral Region (cm ⁻¹)	Pre-processing Technique	Number of Latent Variables	RMSEC (mMol dm ⁻³)		RMSECV (mMol dm ⁻³)		RMSEP (mMol dm ⁻³)	
				Amino Acid	Keto Acid	Amino Acid	Keto Acid	Amino Acid	Keto Acid
1	First Overtones (1540-1812 cm ⁻¹)	Second Derivative Mean Centre	2	40.9	20.3	41.7	20.8	(i) 48.7 (ii) 54.6	(i) 14.9 (ii) 20.7
2	Combination Bands (2032-2400 cm ⁻¹)	Second Derivative Mean Centre	3	38.7	5.7	39.1	5.8	(i) 40.7 (ii) 35.5	(i) 7.8 (ii) 5.6
3	Combined Model (1540-1812 cm ⁻¹ & 2032-2400 cm ⁻¹)	Second Derivative Mean Centre	3	38.1	8.4	38.8	8.6	(i) 41.5 (ii) 34.3	(i) 7.0 (ii) 6.2
4	Mid IR Fingerprint (940-1465 cm ⁻¹)	Mean Centre	5	35.4	7.7	36.3	8.0	(i) 38.7 (ii) 39.5	(i) 9.1 (ii) 15.2
5	Mid IR Fingerprint (940-1465 cm ⁻¹)	First Derivative Mean Centre	4	37.4	7.8	38.3	8.2	(i) 41.5 (ii) 30.1	(i) 13.0 (ii) 9.0
6	Mid IR Genetic Algorithm	Mean Centre	5	35.3	8.7	36.2	9.0	(i) 38.6 (ii) 37.1	(i) 8.6 (ii) 19.4
7	Mid IR Genetic Algorithm	First Derivative Mean Centre	3	33.9	9.3	34.5	9.5	(i) 39.6 (ii) 38.2	(i) 9.5 (ii) 9.7

5.9 Process Application

5.9.1 Middle Infrared

A PLS model for the biotransformation process based on the first derivative spectrum and the important spectral regions identified by the genetic algorithm was determined as being the best model for the biotransformation process. This model had also successfully undergone external validation using two unseen replicates of the biotransformation process.

To further challenge this model, and its robustness, a further replicate of the biotransformation process was carried out off-site at the industrial partner and utilising a different bioreactor system from that used in with the calibration and external validation process replicates.

An additional replicate of the process was carried out on the same scale (total volume of 300 mL) and following the same procedure as before (3.3.3). Since the process was carried out off-site, the samples were stored under refrigerated conditions for transport and allowed to equilibrate back to room temperature before acquiring the infrared spectra (3.4.1 and 3.5.1).

The acquired spectra were converted to the first derivative format and the regions identified by the genetic algorithm extracted from the fingerprint region. Using these spectral features and the best previously constructed PLS model (based on the first derivative spectrum and regions identified by the genetic algorithm), the ABA and KBA concentrations for each of the samples were predicted. These values were then compared with the values obtained from the traditional off-line HPLC reference method.

Errors of prediction for this dataset were 46.3 mMol dm⁻³ for the ABA and 89.8 mMol dm⁻³ for the KBA analyte. These values suggested that the model was able to

predict the the concentration of ABA in the samples within a reasonable error margin but failed to accurately predict the concentration of KBA.

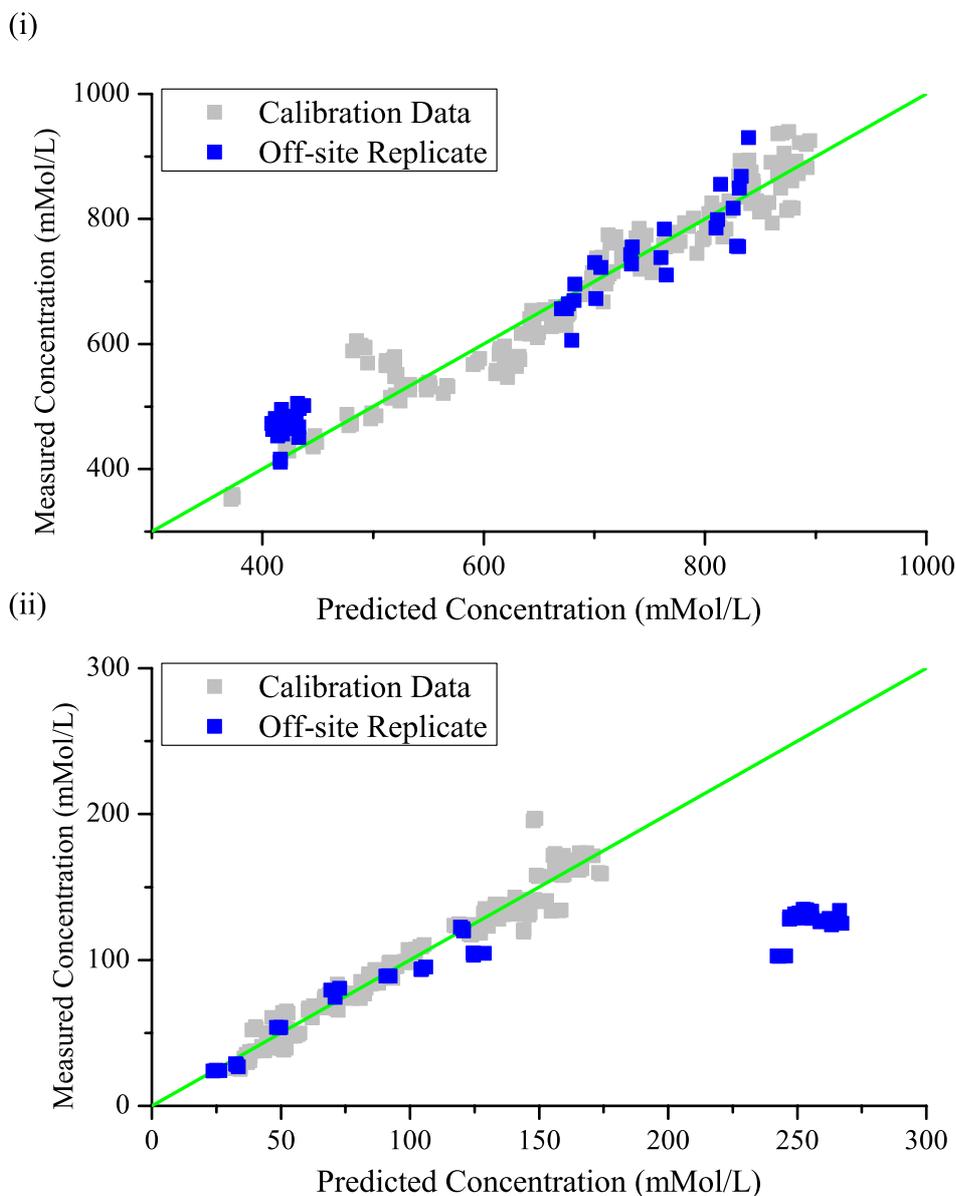


Figure 5-18 - Plots of measured concentration (mMol dm^{-3}) (from the reference analysis) against predicted concentration (mMol dm^{-3}) (from the PLS model) for the off-site replicate of the biotransformation process. (i) Refers to the amino acid (ABA) concentration whilst (ii) refers to the keto acid (KBA) concentrations.

Examination of the plots of measured concentration against predicted concentration revealed some substantial discrepancies for both components (Figure 5-18). In both cases, the first seven samples appear randomly scattered around the central diagonal indicating no bias in the model. However the latter samples (8-16) appear to be wrongly predicting the analyte concentrations. This was particularly pronounced with the keto acid where the model predicted concentrations in the region of 250 mMol dm⁻³ but the measured concentrations were only 120 mMol dm⁻³.

The reference data followed the expected trends and fell within the levels previously identified, which suggested an abnormality with the spectroscopic data. Therefore PCA was carried out on the spectra from the external process replicate and a plot of PC2 scores plotted against PC1 scores constructed (Figure 5-19). Within this plot the latter samples (samples 8 through to 16) of the off-site replicate appeared to occupy a very different spatial region compared with the initial samples and the spectra of the calibration samples. Since the PCA was carried out solely on the spectroscopic data this confirmed that from a spectroscopic perspective something had occurred with the latter samples not evident in the earlier samples or the calibration dataset.

Reference data consistent with the trends observed in previous process replicates and the PCA carried out on the spectra suggested an issue with the spectrometer or spectroscopic data. Temperature effects could be put forward as a possible explanation for the variation in acquired spectra. Differences in temperature due to a longer equilibrium time or localised heat source may have influenced the spectra in some manner. Variations in the infrared spectra due to temperature are a known issue and have previously been reported as affecting the quality of the acquired spectra.⁶⁰ Ideally the spectra of these samples should have been re-acquired and the concentrations predicted from these new spectra. However due to insufficient sample volume remaining following the initial spectral acquisitions and off-line reference analysis it was not possible to re-acquire the mid infrared spectra for these samples.

The constructed model was able to predict the concentrations of ABA in the off-site replicate of the biotransformation process in-line with the errors of calibration and

cross validation over the duration of the process. Examination of the plots of measured concentration against predicted concentration suggested there was a slight bias towards over predicting the ABA content in the latter samples from the biotransformation (Figure 5-18).

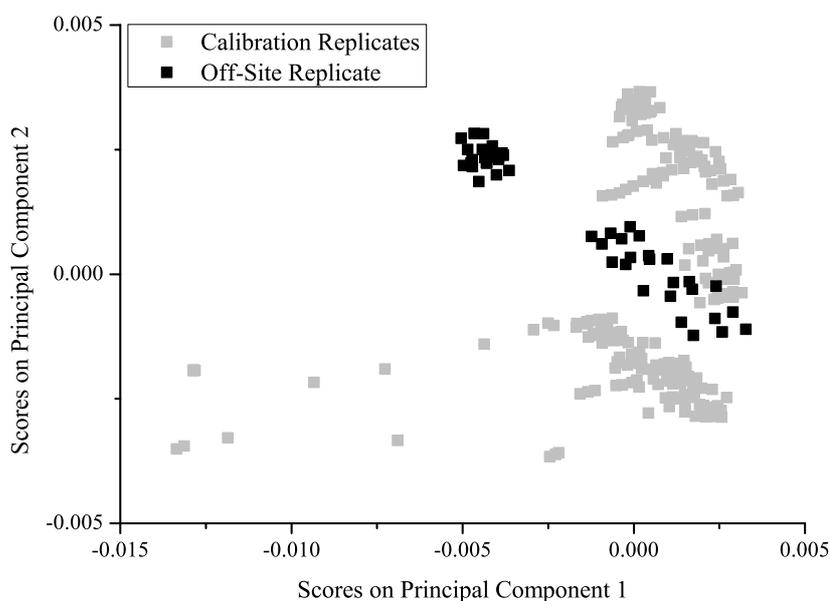


Figure 5-19 – Scores associated with PC 2 plotted against PC1 scores for the mid infrared calibration dataset and the off-site process replicate.

In terms of the keto acid content the model successfully quantified the concentrations of KBA in the first 7 samples of the biotransformation process but failed to correctly predict the concentrations of the remaining samples. Upon examination of the plot of measured concentration against predicted there appeared to be a bias towards over prediction of KBA content in the latter samples (Figure 5-18). Investigation of the spectroscopic data using PCA suggested some spectral variation between the samples obtained during the early stages of the process and the later samples. Differences in the obtained spectra may offer some explanation as to why the model failed to accurately predict the concentrations of the latter samples from the off-site process replicate.

5.10 Conclusions

Based on the results presented for this particular biotransformation it appears that both the near and mid infrared spectroscopic regions could potentially be used to monitor the process.

Within the near infrared region calibration models were constructed based on the first overtones region, combination bands region and a combination of both these spectral regions. Reasonable errors of calibration, cross validation and prediction were obtained for the near infrared models. Evaluation of the regression coefficients associated with these models confirmed subtle differences suggesting the co-linearity issue within the system had been addressed.

Mid infrared models for the system were also constructed with reasonable errors of calibration, cross validation and prediction obtained. Examination of the regression coefficients once again suggested that the measures taken adequately addressed the co-linearity within the system. The best model, as determined from the errors of prediction for the unseen validation datasets, was based on the first derivative mid infrared spectra using the regions identified by a genetic algorithm.

The model was challenged further by presenting samples from a biotransformation process carried out off-site and attempting to predict the ABA and KBA concentrations. Initially the early samples were successful, however the latter samples were not successfully quantified possibly due to a difference in the acquired spectra due to temperature variations or another instrumental or environmental factor.

6. Alanine De-racemisation

6.1 Process Overview

The biotransformation process investigated was the selective de-racemisation of a racemic mixture of alanine using the DAAO enzyme expressed within a modified *Pichia pastoris* cell (Figure 6-1). Details of the procedures and conditions employed during the biotransformation process are given in section 3.3.4.

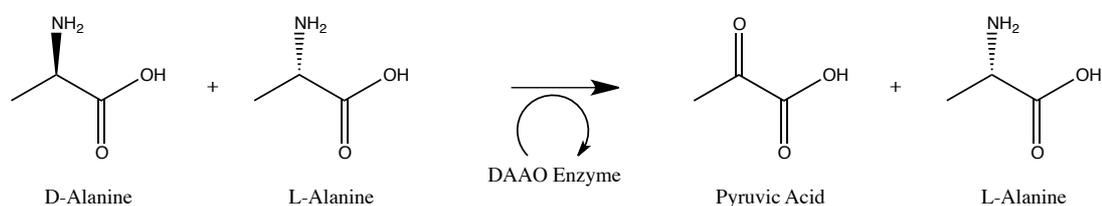


Figure 6-1 - De-racemisation of D/L-alanine using the DAAO enzyme contained within a *Pichia pastoris* host generating L-alanine and pyruvic acid.

Utilisation of mid infrared spectroscopy *in-situ* has been relatively restricted due to the lack of optical fibres that allow for the transmission of radiation in this region between the sample interface and spectrometer. Traditional silica based optical fibres exhibit attenuation of light at wavenumbers below approximately 3333 cm^{-1} (wavelengths greater than $3\text{ }\mu\text{m}$ / 3000nm).⁶⁵ Previously optical fibres based on chalcogenide have been employed with *in-situ* probes. These exhibited good transmission of light with wavenumbers between approximately 550 cm^{-1} and 1000 cm^{-1} ($10\text{-}18\text{ }\mu\text{m}$). Many biological based systems exhibit spectral features, particularly from the amine groups of amino acids, between 1000 cm^{-1} and 1500 cm^{-1} ($7\text{-}10\text{ }\mu\text{m}$).¹⁴³ Therefore even with chalcogenide based optical fibres access to the information rich fingerprint region was not possible, thereby hindering the *in-situ* application of this technique to biological systems.

With the development of silver halide based optical fibres, which allowed for the transmission of light between 550 cm^{-1} to 3333 cm^{-1} ($3\text{-}18\text{ }\mu\text{m}$), the application of *in-situ* probe based mid infrared spectroscopy for many biological based systems became theoretically possible.¹⁴³

Since this process was being investigated using an *in-situ* probe system it was necessary to accommodate the probe within the bioreactor setup. Orientation of the probe should be in a location where the sample it is in contact with is homogeneous, and therefore representative of the sample matrix as a whole. In this particular bioreactor system, the probe was mounted in a vacant port in the head plate alongside the other, more traditional probes for process monitoring such as pH and dO_2 (Figure 6-2). At this location the efficient mixing, as a result of the bioreactor design, ensured the probe was in contact with a representative sample.

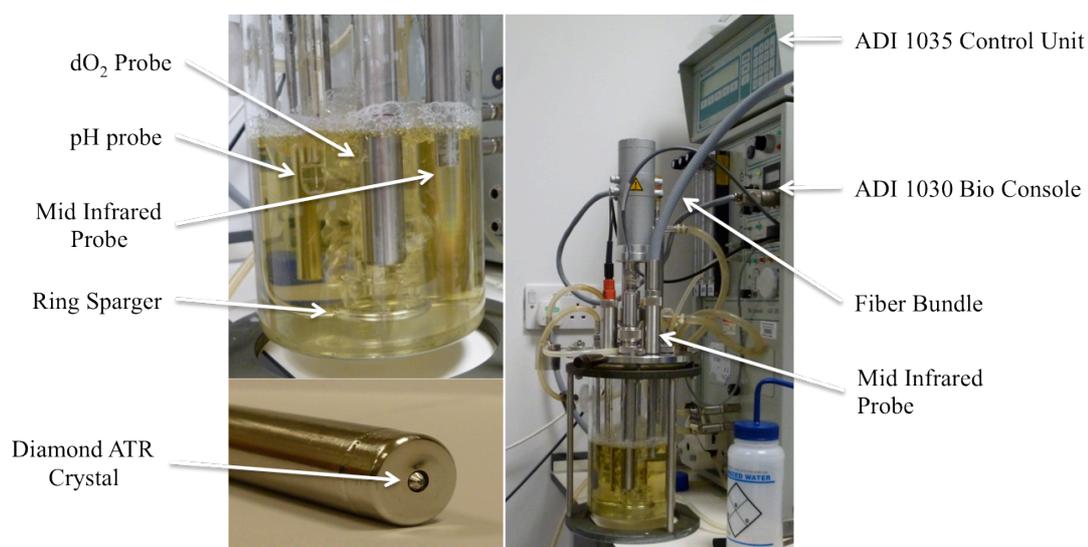


Figure 6-2 - Setup of bioreactor system illustrating the location of the mid infrared probe during the biotransformation process and the probe design with diamond ATR crystal.

The picture of the location of the probe *in-situ* (Figure 6-2) illustrates one of the major challenges that exist when attempting to utilise infrared spectroscopy to monitor biological processes in these environments. By design, the bioreactors ensure efficient mixing and gas dispersion¹³⁷ which creates challenges for

spectroscopic measurement of the sample. Interference effects due to air bubbles were noted as a key challenge in the application of *in-situ* near infrared spectroscopy to such processes.^{20, 32} Despite the mid infrared probe clearly being exposed to this environment the small path-length offered by the ATR crystal in combination with other instrumental parameters was expected to allow suitable spectra to be acquired in this environment. This was based on the assumption that the effects of air bubbles would be reduced in a manner similar to the reduced effects of water signals observed with ATR technology.⁵⁴

6.2 Aim & Objective

This program of work aimed to bring together the findings established using the at-line measurement approaches and apply these to investigate the potential of applying infrared spectroscopy to monitor the process *in-situ*. Other de-racemisation biotransformation processes investigated suggested that whilst in some cases it was possible to follow the process with near infrared spectroscopy, the mid infrared region proved more successful. Whilst the at-line models constructed to date represented a substantial improvement over currently employed methods for process monitoring, the ability to make true *in-situ* near real-time measurements would be the ideal scenario.

Using an *in-situ* diamond ATR probe, the objective was to construct and externally validate a model for this particular alanine de-racemisation biotransformation process. The contributors to any constructed model should also be examined to confirm that each analyte of interest was being independently quantified. Success in this context would demonstrate the feasibility of applying this technique to these challenging biological systems.

6.2.1 Novelty

This work demonstrated the ability to construct a quantitative model for this particular biotransformation process using an *in-situ* mid infrared system. Due to

restrictions on the optical fibres required by these systems their widespread application is still in its infancy compared with the vast array of reported application of *in-situ* near infrared. In this respect, this application represents the application of this technique to a novel bioprocess.

In common with the other chapters, the approach taken to break the stoichiometric is relatively novel in that it is a more structured approach than has previously been employed.

6.3 Analytical Reference Methods

As with all the other processes investigated, it was necessary to have a robust, validated analytical method in place before carrying out any replicates of the biotransformation process. All off-line samples were analysed in triplicate, however since the spectra were acquired *in-situ* only a single spectrum corresponded to each sample. In this case the quoted concentrations that corresponded with a particular sample/spectrum were the mean values from the HPLC triplicate injections.

6.3.1 Quantification of Alanine

Separation and quantification of the alanine content in the biotransformation samples was achieved using the gradient elution method and pre-column derivatisation procedure described (3.2.1.1 & 3.2.1.2.2).

Quantification of alanine using this analytical method, in conjunction with a second analyte of interest, had already been carried out (8.3.2.1). The error of quantification for alanine using this method was determined as being $\pm 0.20 \text{ mMol dm}^{-3}$. The resulting chromatography returned a peak tailing factor of 1.03 and a peak asymmetry factor of 1.08 both of which were acceptable for the current application. Details on the determination of this value are contained within Appendix II.

6.3.2 Quantification of Organic Acids

Two organic acids were produced during the biotransformation process: pyruvic acid and acetic acid. The conditions employed for the separation and quantification of these two components were those detailed in the organic acid method (3.2.2.3).

An assessment of the error associated with this method was based on the analysis of a series of standards at known concentration of these two analytes to construct a calibration curve and assess repeatability of the method (Figure 6-3).

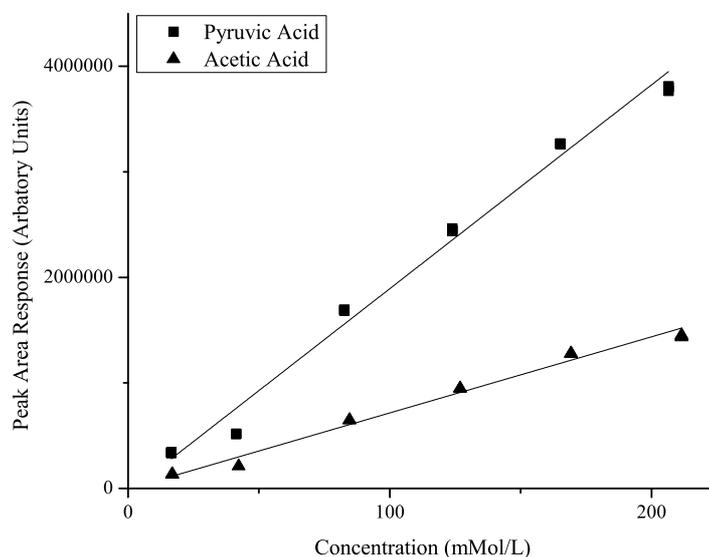


Figure 6-3 - Calibration curves for the quantification of pyruvic acid and acetic acid using the developed organic acids HPLC method.

Adopting the procedure described (3.2.2.4) the error associated with the quantification of pyruvic acid was estimated as being $\pm 7.60 \text{ mMol dm}^{-3}$, and acetic acid being $\pm 8.31 \text{ mMol dm}^{-3}$ using the instrumental conditions described (3.2.2.3) (Appendix II).

The pyruvic acid peak returned an asymmetry factor of 0.71 and tailing factor of 0.86. These values suggested the peak was fronting slightly. Acetic acid peaks had an

asymmetry factor of 1.5 with a tailing factor of 1.17, values which suggested a slight tailing of the peak. In this case the peaks were not as symmetrical as would have been preferred however they were sufficient for the purpose intended.

6.4 Initial Feasibility Study

In common with the other biotransformation processes investigated prior to carrying out any replicates of the process the infrared spectrum of a pure sample of the key components was acquired. These spectra were utilised to identify which functional groups gave rise to each signal but also to confirm whether there were sufficient features in the spectra to allow the construction of independent models.

Since this particular system involved the application of a mid infrared probe *in-situ* other system parameters, such as the spectral resolution and number of co-added scans, also required optimisation prior to any process replicates being carried out.

6.4.1 Pure Components

Pure component solutions of the three key analytes of interest were prepared at concentrations representative of those expected during a typical replicate of the biotransformation process. Alanine was prepared at a concentration of 1000 mMol dm⁻³ and both pyruvic acid and acetic acid at a concentration of 200 mMol dm⁻³. The mid infrared spectrum of each solution was acquired using the mid infrared probe and the spectral features compared. At the outset the presence of acetic acid was not expected within the system however following observations discussed in section 6.5.2.1 the presence of acetic acid in the system became important, the pure component spectra are therefore included in this section for completeness.

Previous attempts at using at-line mid infrared spectroscopy with other similar biotransformation processes had suggested the fingerprint region was the most suitable for modelling. In addition the diamond ATR crystal utilised exhibits a cut-

off region between 1900 cm^{-1} and 2200 cm^{-1} , therefore based on these observations only the spectral region below 1900 cm^{-1} was acquired and investigated (Figure 6-4).

Within the spectrum of alanine a number of signals were noted both in the fingerprint region (below 1200 cm^{-1}) and in the section of the diagnostic region before the ATR crystal cut-off point. Within the fingerprint region a small signal was noted at 1100 cm^{-1} with multiple signals observed in the region between 1300 cm^{-1} and 1450 cm^{-1} . A signal observed between 1500 cm^{-1} and 1600 cm^{-1} was attributed to the amine functional group, most likely bending of the NH bonds lowered and broadened due to hydrogen bonding.¹¹⁵

In the absorbance spectrum of the pure component samples it was difficult to identify some of the spectral features attributable to pyruvic acid. A small signal was noted in the 1150 cm^{-1} region, which was most likely due to vibrations of the CH bonds. The other broad signal noted between 1500 cm^{-1} and 1600 cm^{-1} was common to all three analytes of interest and was most likely attributable to the carboxylic acid group, again lowered in frequency as a result of hydrogen bonding.¹¹⁵

A number of signals were observed in this region of interest for acetic acid as well. A strong, unique signal was observed between 1200 cm^{-1} and 1300 cm^{-1} as well as a unique shoulder on the signal from the carboxylic acid between 1700 cm^{-1} and 1750 cm^{-1} .

Since some spectral features appeared weak in the raw absorbance spectra these were converted to the first derivative format in an attempt to enhance these spectral features (Figure 6-4).¹⁴⁴ On moving to the first derivative spectra there was a clear enhancement in the spectral features of all three analytes. Features arising from pyruvic acid were clearly enhanced with a clear signal now being observed between 1150 cm^{-1} to 1200 cm^{-1} .

Based on the pure component spectra it was concluded that to enhance spectral features of the analytes the first derivative of the absorbance spectra should be

utilised in the modelling process. It was also clear that there were a number of features observed, some unique and some overlapping, between the various analytes of interest that could allow a multivariate model for the biotransformation process to be constructed.

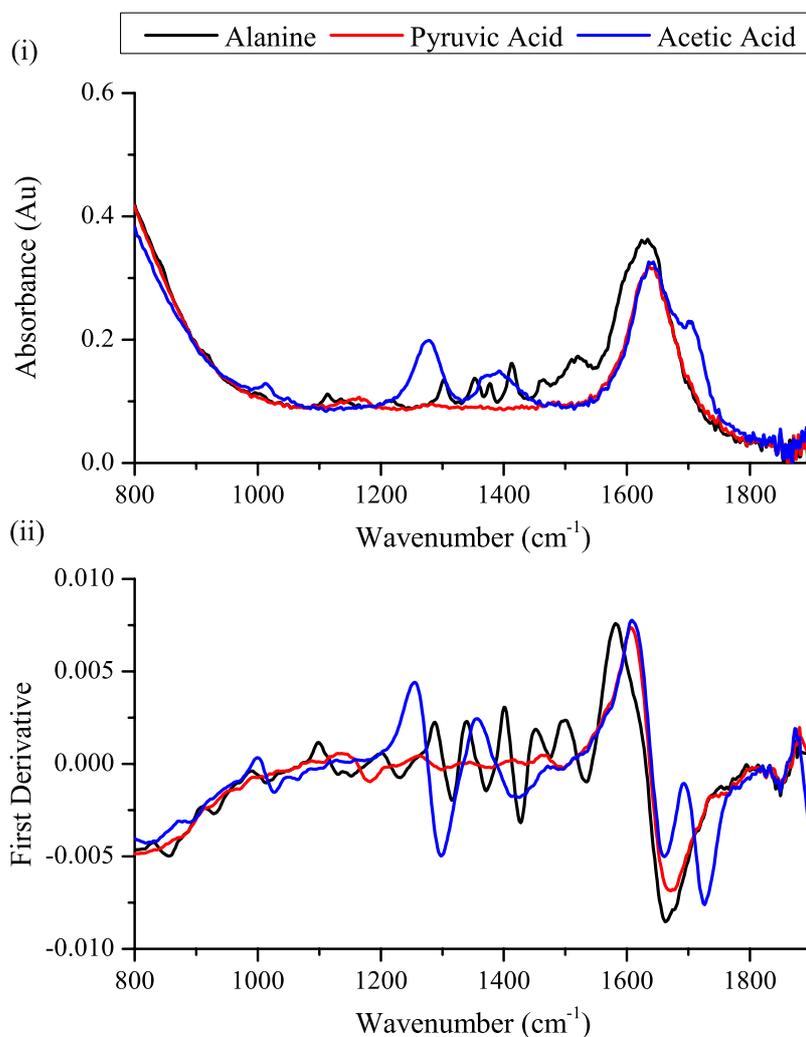


Figure 6-4 - Pure component absorbance and first derivative mid infrared spectra of the three key analytes of interest in the biotransformation process acquired at-line using the mid infrared ATR probe system. Due to absorbance by the diamond ATR crystal only the region between 800 cm^{-1} and 1900 cm^{-1} was investigated.

6.4.2 Parameter Optimisation

Other parameters such as the wavenumber resolution of the spectra, and the number of co-added scans also required optimisation for the *in-situ* system. Initially these parameters were set the same as those used for the at-line sample acquisition (2 cm⁻¹ resolution and thirty two co-added scans) and improved from there.

With the detector system operating at maximum gain to attain the required transmittance through the ATR probe (>20 % throughput) the resulting spectra contained a lot of noise. To reduce this noise (increase the signal to noise ratio) the number of scans averaged to generate the final spectrum was increased from 32 to either 64 or 128 scans of the sample material.¹⁷ Increasing the number of scans increased the overall time taken for the spectra to be acquired however there was an improvement in the quality of the spectra when averaging a greater number of scans.

The wavenumber resolution utilised also influenced the quality of the resulting spectra. Samples of the biotransformation solution were acquired at 2, 4, 8 and 16 cm⁻¹ resolutions and compared to determine which was the optimum resolution for acquiring the spectra. The absorbance spectra at each resolution were then plotted to determine the optimum resolution (Figure 6-5).

At the highest resolution (2 cm⁻¹) the acquired spectrum appeared to be excessively noisy particularly in the region above 1400 cm⁻¹. This was the resolution that had previously been employed when acquiring the at-line mid infrared spectra. Decreasing the resolution slightly to 4 cm⁻¹ decreased the noise observed in the spectra substantially whilst still retaining the spectral features of interest. Further decreasing the resolution to 8 cm⁻¹ again resulted in a reduction in the background noise with most of the key features still being identifiable. When the resolution was further reduced to 16 cm⁻¹ the acquired spectra did not show any obvious signs of noise however some of the spectral features were beginning to disappear, of particular note was the absence of the signal at 1100 cm⁻¹ that corresponded with the alanine.

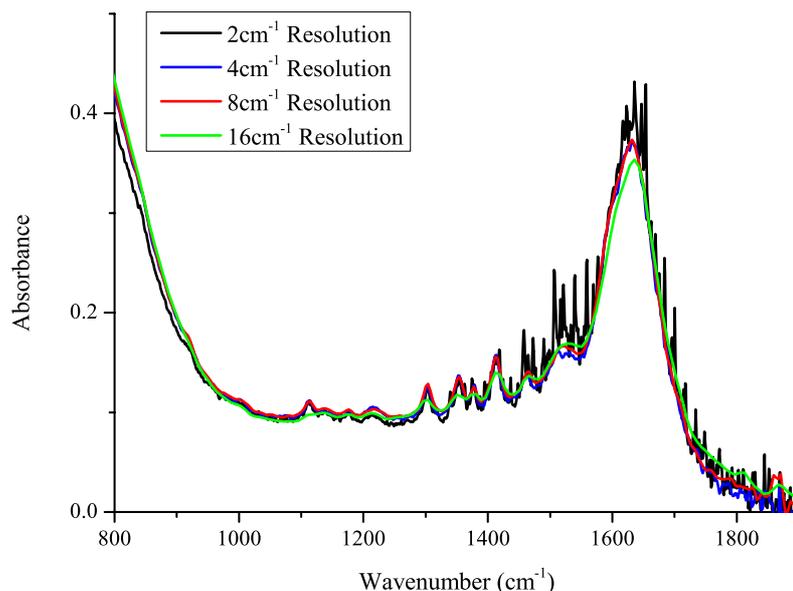


Figure 6-5 - Mid infrared spectra of biotransformation samples obtained at various resolutions to determine the optimum parameters for the spectral acquisition.

Based on these observations the spectra were acquired at a resolution of 4 cm^{-1} as this appeared to provide the maximum wavenumber resolution without the spectra becoming dominated by background noise. During the biotransformation process each acquired spectrum was the average of 128 co-added scans. Despite this resulting in the spectral acquisition process taking in approximately 6 minutes to complete, it resulted in the highest quality of spectra for use in the modelling process. The rate of reaction for this particular process was also unlikely to result in substantial changes in concentration causing an averaging effect over time period the spectra were acquired.

6.4.3 Summary

The observations from the pure component samples determined that for the probe and spectrometer system available the best quality spectra were obtained with a resolution of 4 cm^{-1} and was the result of 128 co-added scans. To enhance some of

the spectral features, particularly those resulting from pyruvic acid, the acquired spectra were converted to the first derivative format for use in the modelling process.

6.5 Biotransformation Replicates

In order to construct a calibration model for the biotransformation process a number of replicates were carried out. Ruckebusch *et al.* (2002) suggested a minimum of five process replicates was required for the construction of a robust calibration model.⁸³ Therefore as a minimum, six replicates of the biotransformation process were required to allow for both calibration and external validation.

Due to the nature of the process under investigation a stoichiometric link was expected to exist between alanine and pyruvic acid. Although the feasibility study had taken into account an additional organic acid, acetic acid, at this point it was not expected to be produced and so was not considered to be stoichiometrically linked with alanine or pyruvic acid.

To break this co-linearity within the system, each replicate of the biotransformation process was carried out with a different initial starting concentration of alanine and pyruvic acid as had been previously reported and adopted.¹³⁶

6.5.1 Experimental Design

Initial concentrations of alanine and pyruvic acid, in the six replicates of the biotransformation process, were varied according to a mixed level (alanine had three levels, pyruvic acid only two) experimental design (Table 22). The experimental design was generated using the Design Expert software application [Version 8] (Stat-Ease Inc., Minneapolis, USA). At the start of the biotransformation process no pyruvic acid was expected to be present. By introducing a spike, not only was the co-linearity addressed, there was a residual concentration of pyruvic acid which would result in a more accurate quantification since there should be no samples below the analytical methods limits of detection.

Table 22 - Experimental design used to determine the initial concentrations of alanine and pyruvic acid added to each replicate of the biotransformation process to ensure the stoichiometric linkage was broken.

Biotransformation Replicate	Alanine Concentration (mMol dm⁻³)	Pyruvic Acid Concentration (mMol dm⁻³)
1	1000	200
2	1200	200
3	800	100
4	1200	100
5	1000	100
6	800	200

6.5.2 Process Replicates

Six replicates of the biotransformation process were carried out as described (3.3.4) with the initial concentrations of alanine and pyruvic acid spiked according to the experimental design (Table 22). Ideally all six replicates of the process would have been utilised as the calibration dataset, with additional validation replicates also carried out. However due to time constraints on the availability of the instrumentation, the adopted approach allowed for the best calibration model to be constructed based on the imposed variations within the system but also ensuring that the known issue of co-linearity was adequately addressed in the model.

The mid infrared spectra, with a resolution of 4 cm⁻¹ and 128 co-added scans, was collected at ten minute intervals over the course of the biotransformation process (3.4.2). Samples were removed at regular intervals over the course of the process, heated at 100 °C to prevent further conversion, and the concentrations of the key analytes determined.

6.5.2.1 Initial Observations

When the samples from the first replicate of the biotransformation process carried out were quantified for their alanine and pyruvic acid concentration some unexpected values were observed.

The total concentration of alanine in the samples decreased as the biotransformation process progressed as expected.

Initially a spike of approximately 200 mMol dm^{-3} pyruvic acid was introduced into the system. As the reaction progressed this was expected to increase however the opposite effect was observed. From the initial concentration (already substantially less than the intended spiked level) the concentration decreased before stabilising in the $15\text{-}20 \text{ mMol dm}^{-3}$ region. These results were contrary to expectations since the alanine concentration profile still suggested conversion of D-alanine to pyruvic acid.

A possible explanation for this decrease in concentration of pyruvic acid is that it was being utilised by the *Pichia pastoris* cells. Pyruvic acid can be converted directly to acetyl-CoA before moving into the citric acid cycle (Figure 6-6). Alternatively, pyruvic acid may undergo a carbon dioxide fixation step to produce oxaloacetate, an intermediate compound found in the citric acid cycle carried out by the organism for the generation of energy (NADH & FADH₂).¹⁵

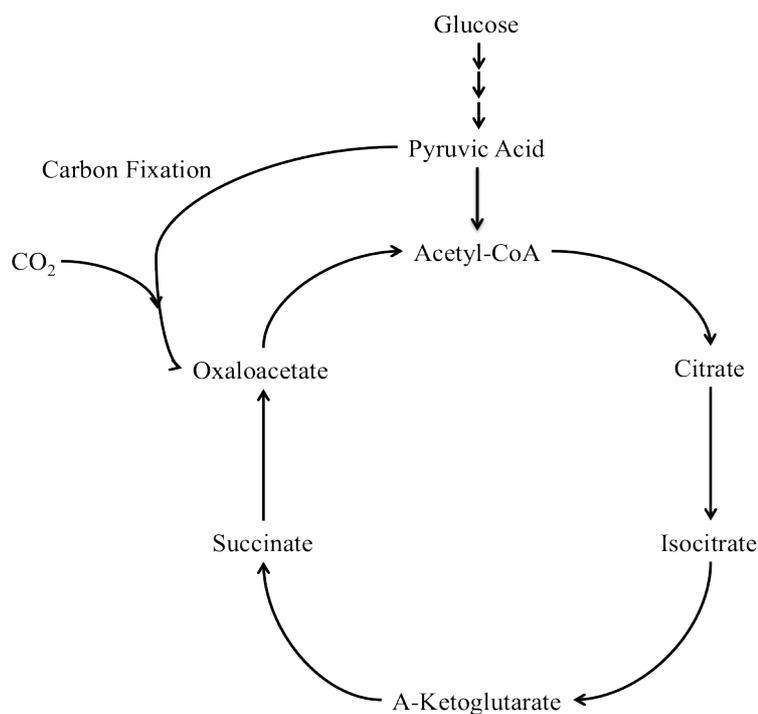


Figure 6-6 - TCA cycle adapted from Prescott *et al.* illustrating the routes by which pyruvic acid can be utilised either by conversion to acetyl-CoA or carbon fixation into the TCA cycle for energy generation.

However, when quantifying the organic acid content of the samples using the HPLC method (3.2.2.3) the appearance of an additional peak on the chromatograms was observed as the biotransformation process progressed. The proposed mechanism did not explain the appearance of this additional peak. In order for carbon fixation to occur it requires the presence of the cofactor biotin, which had not been added to the system.

This would therefore suggest it was unlikely that the pyruvic acid was being utilised by the cells; either by carbon fixation due to the absence of biotin or in the TCA cycle due to the appearance of an additional peak. It was more likely that pyruvic acid was undergoing conversion to another organic acid that was being detected by the HPLC method.

The action of the DAAO enzymes on the substrate, oxidising the D-enantiomer to the corresponding keto acid, was facilitated by the reduction of FAD (2.1). Regeneration of FAD with oxygen, supplied via the air supply, resulted in the generation of hydrogen peroxide as a by-product. In the biotransformation involving the conversion of TBG to TMP catalase was utilised as a scavenger for the hydrogen peroxide, however in the present system no catalase, or alternative, was present to remove this by-product.

The pyruvic acid produced by the biotransformation process can be oxidised to acetic acid by the hydrogen peroxide by-product produced during the biotransformation process (Figure 6-7).¹⁴⁵ To confirm the unidentified peak noted in the samples was acetic acid, a pure sample was analysed using the HPLC method and the retention times compared. Both the retention time of the acetic acid standard and the unidentified peak in the actual process samples matched suggesting that the pyruvic acid was being converted to acetic acid over the course of the biotransformation process.

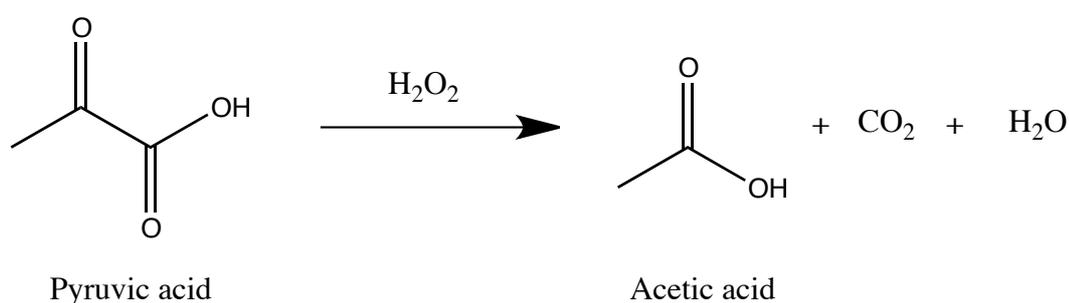


Figure 6-7 - Oxidation of pyruvic acid to acetic acid, water and carbon dioxide by hydrogen peroxide.¹⁴⁵

The concentration of acetic acid was therefore expected to steadily rise as the biotransformation process progressed, however in some replicates of the process there did not appear to be any accumulation of acetic acid in the system. It was possible for some fungi, bacteria and algae to utilise acetic acid as a carbon source for growth.¹⁵ If yeasts (specifically *Pichia pastoris*) were also able to utilise acetic acid as a carbon source this would explain the lack of accumulation observed.

Based on these observations from the first biotransformation process the analytical method employed (3.2.2.3) was used to quantify both pyruvic acid and acetic acid in the samples. The process conditions were not altered to incorporate a hydrogen peroxide scavenger such as catalase as it was unlikely this would be employed in an industrial context. Pyruvic acid, and acetic acid, in this process were undesired by-products that were of interest from a modelling perspective but not from an industrial perspective. The process carried out at the lab scale should remain as consistent with the industrial process as possible, these analytes still remained of interest from a spectroscopic viewpoint however the main focus was therefore to focus on monitoring the alanine content in the system.

6.5.2.2 Data Interpolation

The six replicates of the biotransformation process, with the initial concentrations of racemic alanine and pyruvic acid varied according to the experimental design, were carried out. Infrared spectra of the system were acquired at ten-minute intervals over the duration of the process. Samples were removed from the system at regular intervals, generally hourly, during the initial phases of the process with the time period increasing as the reaction progressed and conversion slowed. This meant that the total number of spectra acquired during the biotransformation substantially outnumbered the corresponding reference measurements. To compensate for this it was necessary to interpolate the missing data points using the available reference measurements.

6.5.2.2.1 Linear Interpolation

Average alanine concentrations for each sample removed during the biotransformation process were plotted against the time point at which the sample was taken. Each point was joined to the previous with a linear line and the alanine concentration for each spectrum acquired estimated.

This approach to interpolation of the data was not particularly accurate. The error associated with the reference analysis method meant there was some variance in the quoted average concentration values. Since the linear line drawn was between two data points, and not a best fit through all the data points, any erroneous values would then cause the missing values to be incorrectly quantified (Figure 6-8). This would result in the predicted concentrations mimicking any erroneous trends in the data and not an average of the results. In addition this approach assumes that the overall biotransformation process obeys a linear relationship, which was an incorrect assumption clearly illustrated in Figure 6-8.

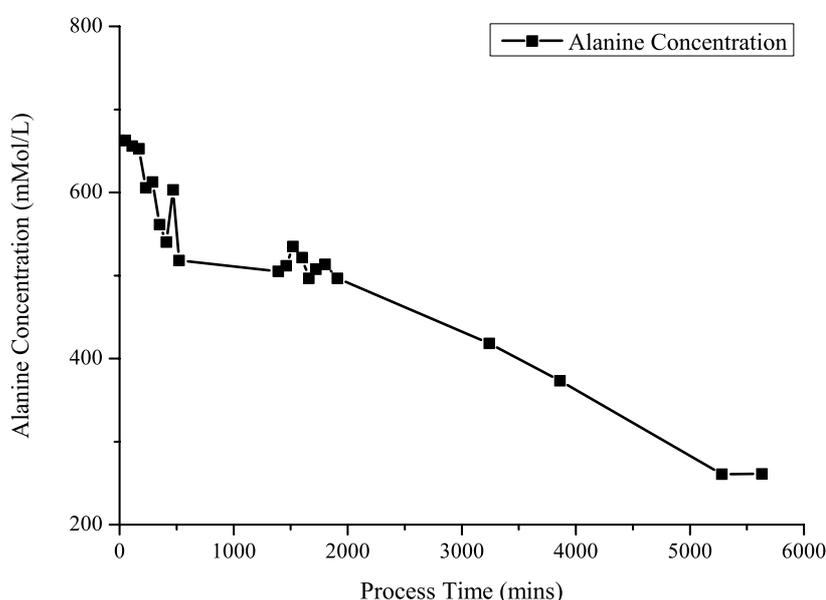


Figure 6-8 – Mean alanine concentration plotted against process time. Interpolation of the unknown values using a linear approach where a linear line connected the data points did not accurately represent the missing points. From the shape of the plot it was also clear that a linear trend was not the best fit for this dataset.

6.5.2.2.2 Exponential Decay

To try and better represent the biotransformation process and its progression, an exponential decay curve was fitted through the reference data points. It was hoped

this would better represent the initial stages of the biotransformation where conversion of D-alanine occurred faster before slowing towards the end of the process.

Fitting an exponential decay through the data points (Figure 6-9) suggested a better fit through the data initially however, once past the initial phase of the decay the curve failed to appropriately fit the data, with suggested concentrations in the middle phase of the biotransformation being far removed from the measured concentrations.

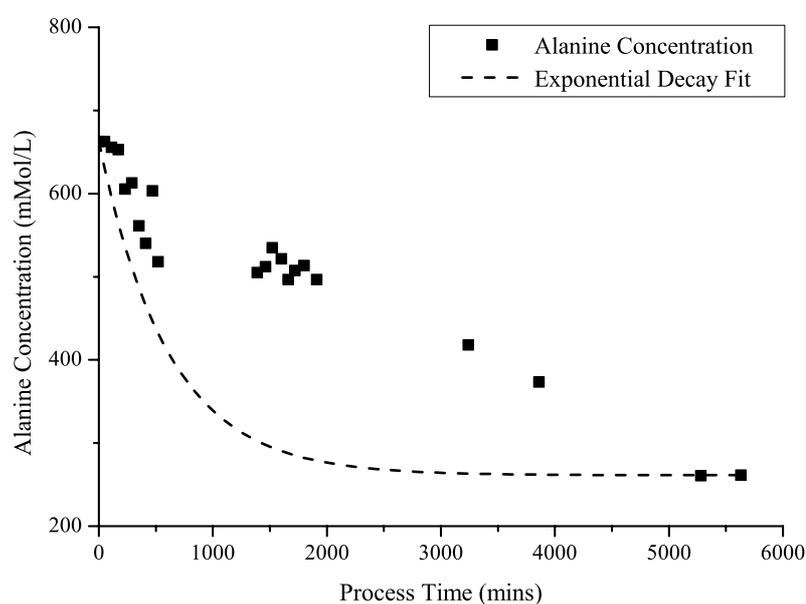


Figure 6-9 – Exponential decay fitted through the reference data points to give a more accurate interpolation of the alanine concentration at the time points for each spectral acquisition.

6.5.2.2.3 Kinetic Interpolation

With little success fitting both linear and exponential decays through the reference data points, an alternative method based on the kinetic approach reported by Trevisan *et al.* (2008) was considered.⁷²

In a simplistic view, the rate at which the biotransformation progressed was independent of the DAAO enzyme and dependent on the concentration of alanine in the system. To determine if the biotransformation process progressed via first order kinetics with respect to alanine, the natural log of concentration was plotted against time (Figure 6-10). A linear line was fitted through these points and evaluated to determine if a linear fit through the data was appropriate. In this case a linear line was a good fit through the data points and so it was experimentally determined the biotransformation process progressed via first order kinetics with respect to alanine concentration.

This process was repeated and the equation of the linear line of best fit for each replicate of the biotransformation determined. Using these equations the estimated alanine concentrations at the time point when each spectrum was acquired were determined and used for the construction of a calibration model.

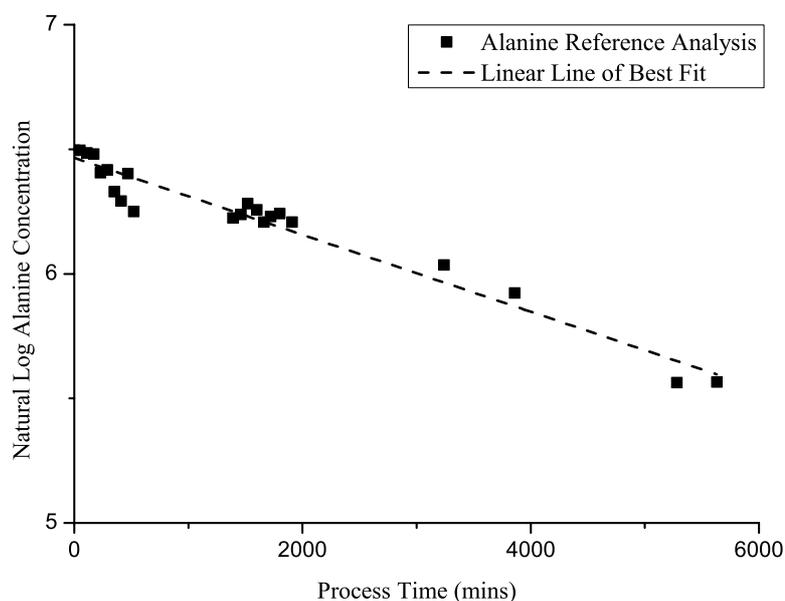


Figure 6-10 - Plot of natural log of alanine concentration against process time. Having determined the process proceeded via first order kinetics, the concentration at the time each spectrum was acquired was estimated based on the equation of the linear line of best fit.

This process of determining the alanine concentration by kinetic fitting through the data was repeated for the six calibration replicates of the biotransformation process. Since an interpolated alanine concentration was now available for each spectrum acquired during the process, a calibration model for the system was constructed.

6.5.2.3 Alanine Model

Having estimated an alanine concentration to correspond with each spectrum acquired during the biotransformation process the six replicates was split into calibration and validation datasets. In order to identify which samples should be included in the calibration dataset a PLS model using the data from all six replicates was constructed and the scores plots examined (Figure 6-11). Previously identification of the datasets in this manner had been achieved using PCA. This approach only takes into account the spectral data whereas the PLS approach considers both the spectral and reference data. Including the reference data in this situation was important since the values were determined using the kinetic fitting approach.

The PLS model consisted of the full spectral region (800 cm^{-1} through 1900 cm^{-1}) in the first derivative format. Both the spectral data and alanine concentrations were mean centered with LOOCV also applied within the model. Examination of the plot of RMSEC/RMSECV values against number of latent variables identified five latent variables as optimum. From the scores plots (Figure 6-11) it was identified that replicates 1,3,4,5 and 6 would be included in the calibration dataset with replicate 2 retained for external validation of the constructed model.

Plots of RMSEC/RMSECV values against the number of latent variables in the model for the calibration data indicated an optimum number of five latent variables. Errors of calibration and cross validation associated with this model were 15.7 mMol dm^{-3} and 15.9 mMol dm^{-3} respectively. These values were better than the errors obtained previously for the amino acid component in similar systems based on the at-line spectra.

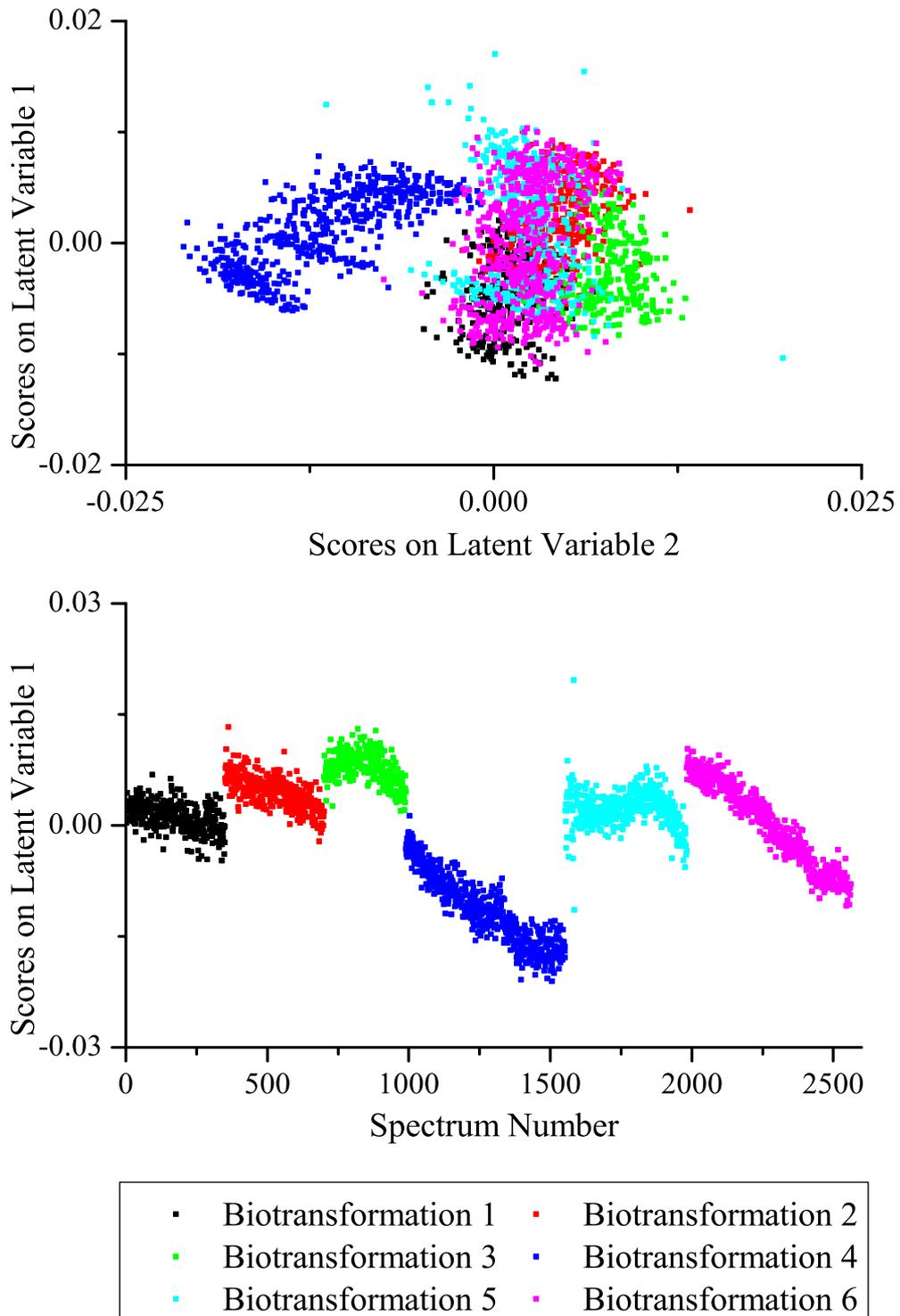


Figure 6-11 – (i) Plot of scores associated with latent variable 1 plotted against scores for latent variable 2 and (ii) scores associated with latent variable 1 plotted against spectrum number to identify the five datasets to be used for calibration and the external validation dataset.

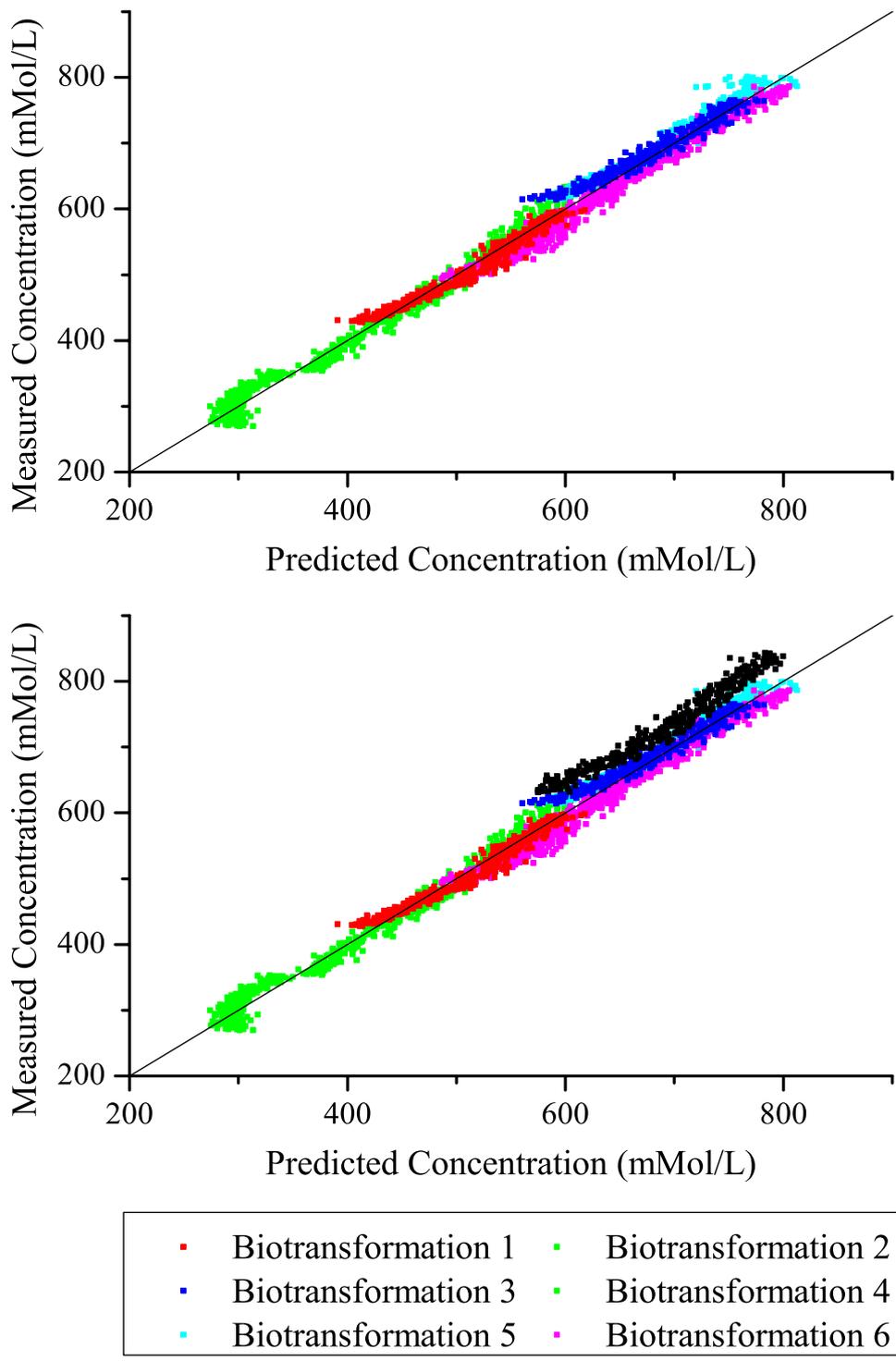


Figure 6-12 - Plots of measured concentration against predicted concentration. (i) Internal validation of PLS model constructed using the full spectral range acquired (800 cm^{-1} through 1900 cm^{-1}). (ii) External validation dataset alongside calibration data.

The measured concentrations, determined by the kinetic fit of the data, were plotted against the predicted concentration, from the internal validation, to assess the robustness of the model and ensure bias in the system was not an issue (Figure 6-12 (i)). Measured values for the calibration dataset were in very good agreement with the predicted values obtained from the constructed model. Points appeared randomly scattered around the central diagonal (1:1) with no obvious bias in the model towards either over or under predicting the concentration of alanine in the system.

Spectral data from the second biotransformation replicate were then input into the constructed model and the alanine concentration of these samples predicted. Predicted concentrations were then compared with those values obtained from the kinetic fit through the HPLC data points. The error of prediction was determined as being 43.4 mMol dm⁻³ with an r² value of 0.96 for this particular dataset. An error of this magnitude was consistent with what had previously been observed with models of similar systems constructed using the at-line spectroscopic techniques. For comparison this was also consistent with the upper error margin reported by Riley *et al.* (1998) for the prediction of alanine using near infrared spectroscopy.³⁸ Examination of the plot of measured concentration plotted against predicted concentration suggested there was a slight tendency towards under predicting the alanine concentration (Figure 6-12 (ii)). Improvements to the model may result if refinement of the spectral regions to eliminate the inclusion of redundant spectral features in the model was carried out.

Closer examination of the first derivative pure component spectra (Figure 6-4) confirmed there were regions where the spectra were common between all analytes or contained no spectral features. The spectral region was therefore reduced to incorporate only the features between 1070 cm⁻¹ and 1555 cm⁻¹. Selecting this region retained the different spectral features noted for the three key analytes but removed the additional noise noted in the spectra as the diamond cut-off region was approached.

The PLS model for the system was re-constructed using this refined spectral region and the alanine concentration for the external validation dataset predicted. Using this reduced region the resultant PLS model consisted of three latent variables and returned errors of calibration and cross validation of $19.3 \text{ mMol dm}^{-3}$ and $19.4 \text{ mMol dm}^{-3}$ respectively. These values were higher than had been observed when using the full spectral region, however the model was based on a smaller number of latent variables. Increasing the number of latent variables in this particular model resulted in over-fitting of the data, indicated by an increase in the difference between RMSEC and RMSECV values. Evaluation of the plot of measured concentration against predicted concentration (from the internal validation procedure) once again illustrated that the calibration data was randomly scattered around the central diagonal suggesting there was no bias in the model (Figure 6-13 (i)).

Prediction of the alanine concentration for the external validation dataset returned an error of prediction of 21 mMol dm^{-3} and an r^2 value of 0.99. This value was in good agreement with the errors of calibration and cross validation. Examination of the plot of measured concentration against predicted concentration suggested there was a slight tendency towards under predicting the alanine concentration. However the values for this unseen dataset appeared to lie within the range covered by the calibration data.

Given that the model was constructed and validated on a modest dataset, this slight under prediction was not a cause for concern and may be a result of the experimental approach taken. Six replicates of the biotransformation process were carried out according to a simplistic experimental design. By removing one replicate from this design the calibration dataset no longer spanned the full experimental region. In counterargument however, since any co-linearity within the system should have been addressed even by removing one replicate, other process replicates will have covered this concentration range. Perhaps, retrospectively, a more appropriate approach would have been to utilise all six experimental design replicates as the calibration dataset with an additional process replicate used to externally validate the model.

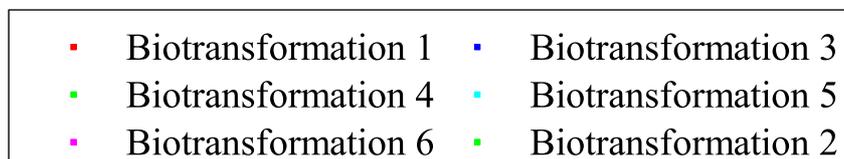
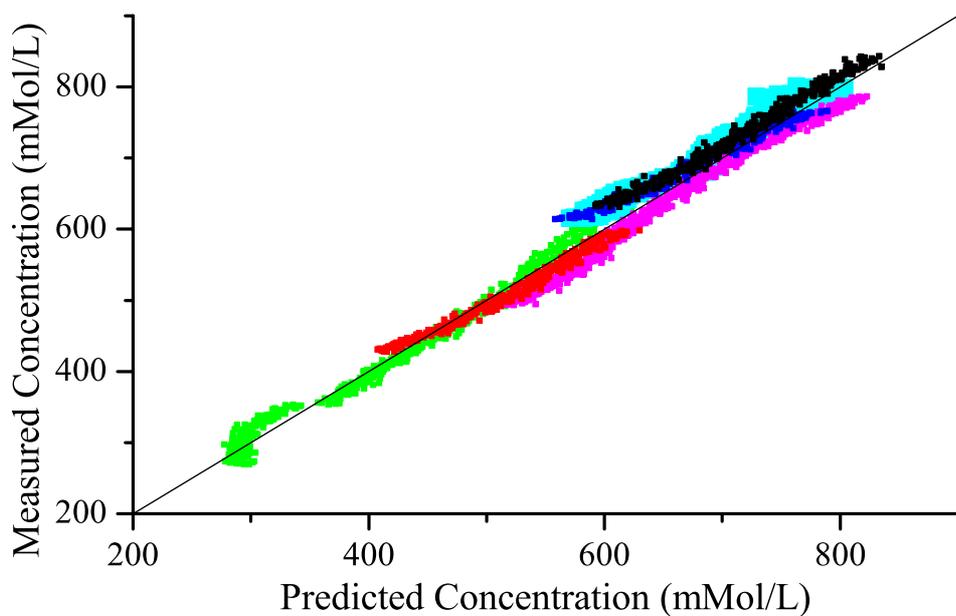
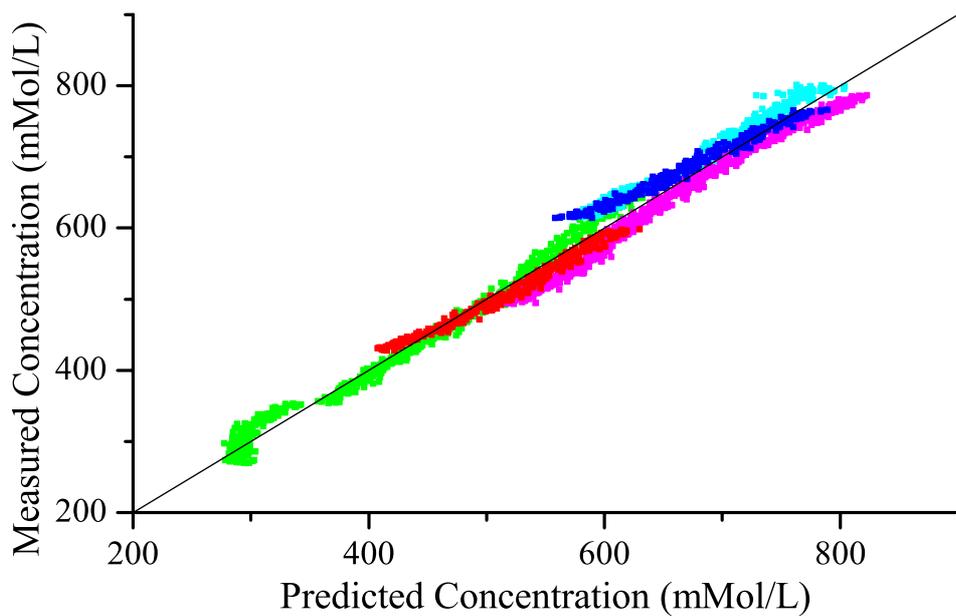


Figure 6-13 – Plots of measured alanine concentration against predicted concentration for PLS model constructed using refined spectroscopic regions. (i) Calibration data with predicted concentrations from internal validation procedure. (ii) External validation dataset shown alongside internal validation data.

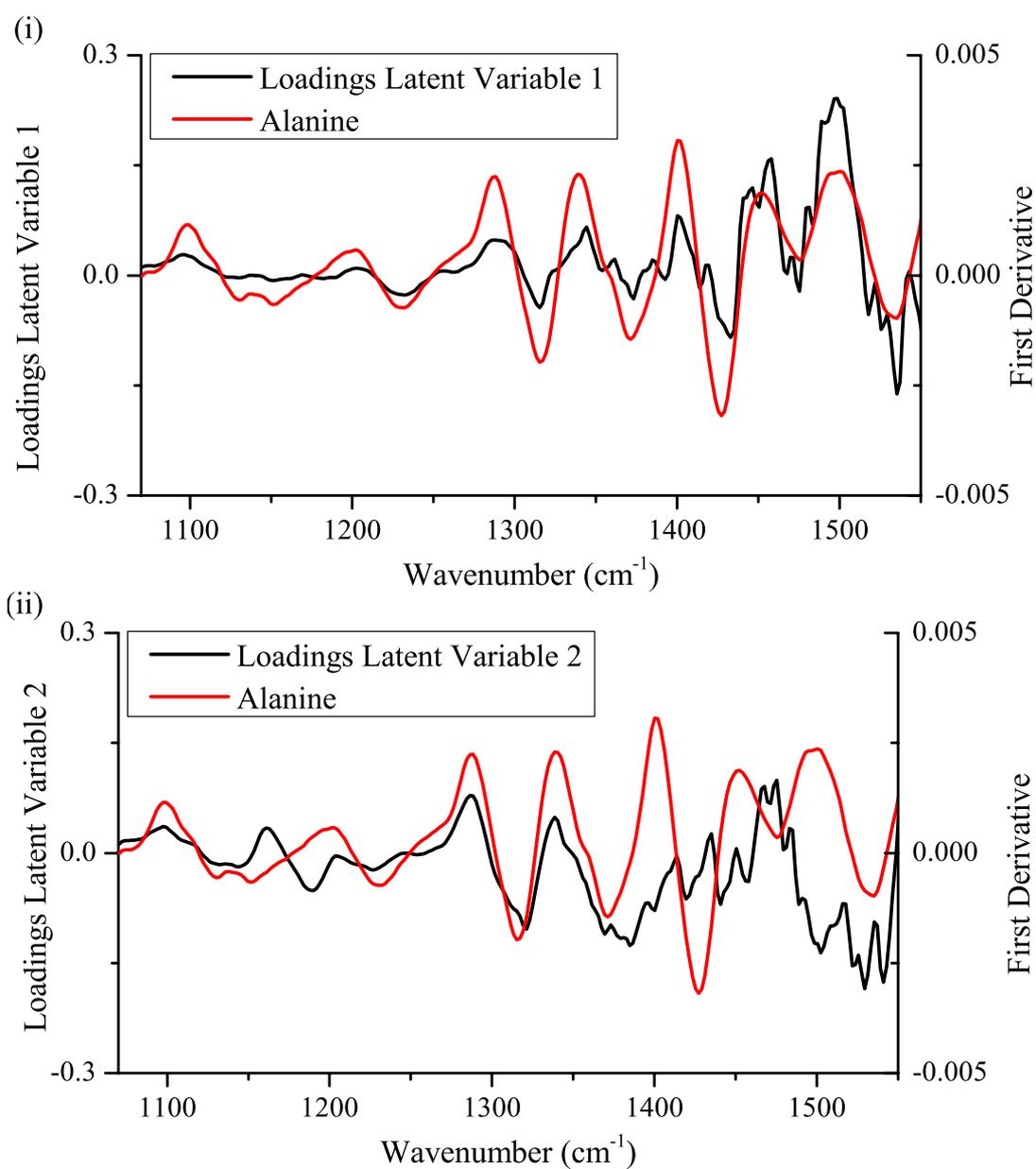


Figure 6-14 – (i) Loadings associated with latent variable one of the constructed model plotted along with the pure component spectrum of alanine and (ii) loadings associated with latent variable two of the model plotted with the pure component spectrum of alanine.

Contributors to the resulting model were also examined to confirm the model was functioning as expected. The loadings plot for the first latent variable in the model suggested a correlation with alanine concentration (Figure 6-14 (i)). Points with a high positive or negative loadings value corresponded with the peaks and troughs in the first derivative pure component alanine spectrum. This held true over the wavelength region (1070 cm^{-1} to 1550 cm^{-1}) used to construct the refined PLS model.

Loadings for the second latent variable only appeared to overlap with the pure component alanine spectrum at a single point in the 1300 cm^{-1} region (Figure 6-14 (ii)). Over the remainder of the spectral region there was little correlation or anti-correlation with the pure component alanine spectra noted.

The regression coefficient obtained for the constructed model was almost an identical match with the spectrum of the pure component alanine (Figure 6-15). This confirmed that the constructed model was quantifying based on spectral features attributed to the alanine.

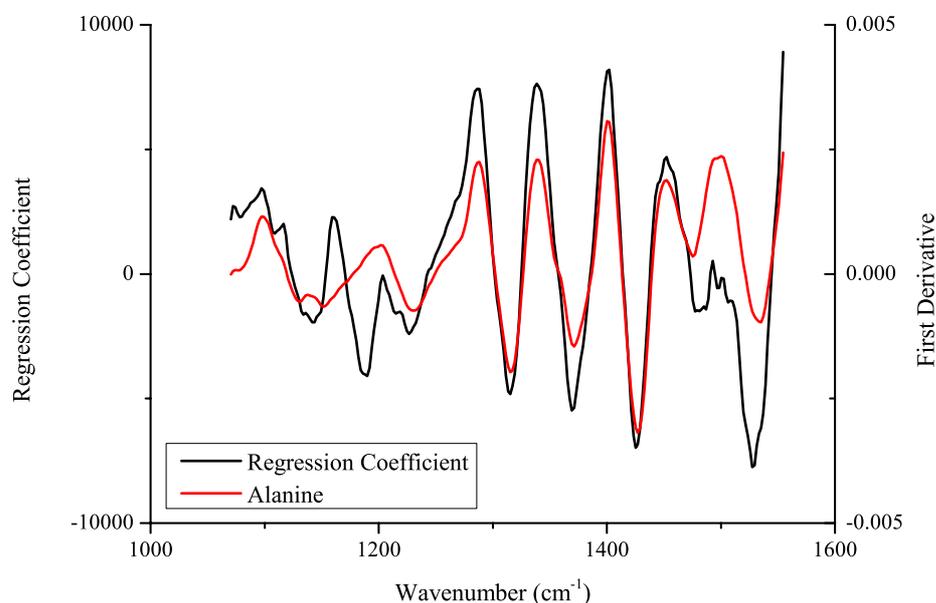


Figure 6-15 - Regression coefficients associated with the constructed PLS model and infrared spectrum of pure component alanine for comparison.

These results have illustrated that it was possible to utilise mid infrared spectroscopy *in-situ* to monitor the alanine concentration of a DAAO de-racemisation biotransformation process. Errors of calibration, cross validation and prediction for the unseen dataset were in good agreement with each other, and were also lower than had been observed with similar systems with the spectra acquired at-line. It would be expected that the more challenging conditions in the reactor would increase the error margins for *in-situ* measurements however, this does not appear to be the case in this situation or in other reported cases.²⁵

There are a number of possible explanations for this improvement in errors of prediction on going from at-line to *in-situ* spectroscopic methods. With the *in-situ* approach only a single spectrum was acquired and corresponded with a particular sample whereas with the at-line approach the spectrum of each sample was acquired in triplicate. This may have introduced subtle spectrometer drift or sample drift, due to effects such as sedimentation, into the resulting model that may influence its predictive ability.

Another factor that may influence these improved error values is the method by which the reference data was calculated. By estimating the reference values for the intermediate points based on the kinetic fit of the data any erroneous values will be less influential and averaged over the course of the process. With an at-line model each sample has a measured corresponding reference value that will impact on the overall model.

Finally the model constructed in this particular case was a PLS1 model, whereas previously PLS2 models have been employed. Since the model was based around a single analyte and not multiple analytes, as had been the case with PLS2 models, it was not unsurprising that better error values were observed in this instance.

6.5.2.1 Organic Acid Model

Having successfully utilised mid infrared spectroscopy to monitor the alanine concentration over the course of a biotransformation process, the above approach was repeated for the organic acid components.

Pyruvic acid was added to the system at varying concentrations to ensure that co-linearity within the system had been addressed. On evaluation of the reference analysis however the concentration of pyruvic acid was noted to fall from the initial levels as discussed previously with the initial observations (6.5.2.1).

The pyruvic acid was converted to acetic acid, which in turn appeared to be utilised by the cellular material. Since there were more spectra recorded than the number of samples taken it was necessary again to try and fit an equation through the sample data to interpolate the organic acid concentrations for each acquired spectrum.

The concentration of pyruvic acid and acetic acid were plotted against time for the biotransformation replicates to identify the most appropriate fit for the data (Figure 6-16). Following the initial decrease in pyruvic acid from the starting spiked concentration, the levels appeared to remain relatively stable over the course of the biotransformation process.

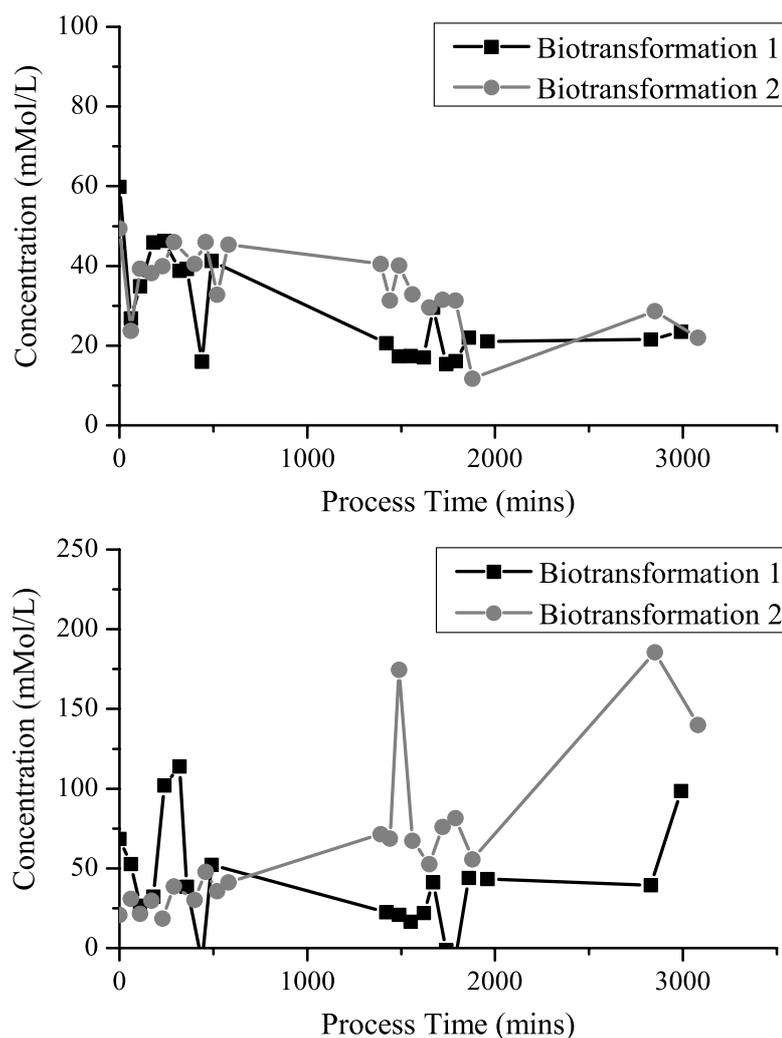


Figure 6-16 - Plots of pyruvic acid and acetic acid concentration against process time for two biotransformation process replicates illustrating (i) the consistency in pyruvic acid concentration during the process and (ii) the consistent concentrations of acetic acid before accumulation towards the end of the process.

Examination of a key spectral feature unique to pyruvic acid at 1245 cm^{-1} confirmed no substantial change during the process. Plotting the first derivative response at this particular wavelength against process time over the duration of the biotransformation process suggested there was very little overall variation observed in this particular

region (Figure 6-17). Variations that were observed appeared more to resemble noise in the system, which was not unexpected since the spectrometer detector was operated at the maximum gain setting.

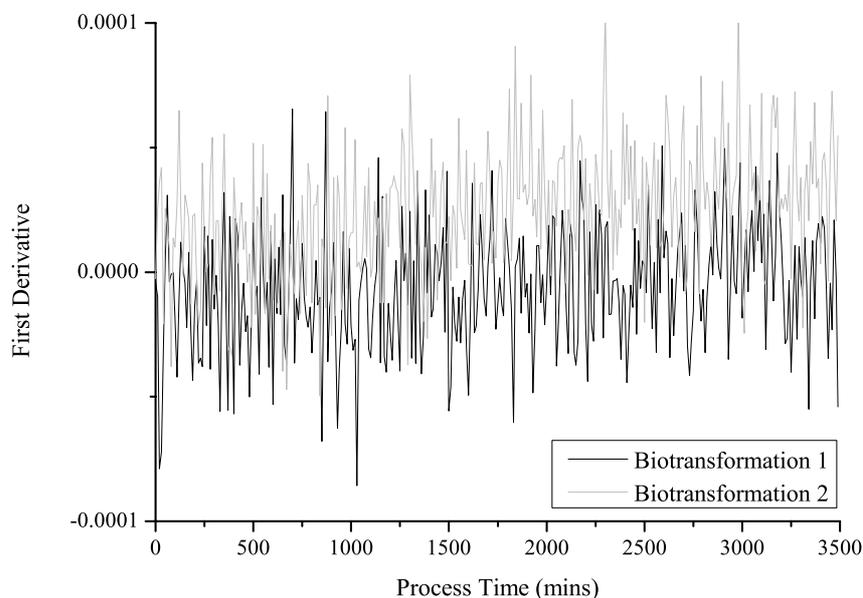


Figure 6-17 - Plots of first derivative value at 1245 cm^{-1} plotted against process time to illustrate the little overall variation observed in the spectral features over the course of the biotransformation process. This confirmed the reference analysis data suggestion that little or no variation in pyruvic acid concentration was observed during the process.

The concentration profile and the spectral data did not appropriately fit an exponential decay trend, nor did it follow first or second order kinetics. Concentrations of pyruvic acid, and acetic acid, appeared to reach and remain at equilibrium during the course of the biotransformation process. Therefore it would be of little value to attempt to construct a model for these analytes since the concentrations are in equilibrium and do not change substantially over the course of the biotransformation process.

6.6 Conclusions

The results obtained have demonstrated that it was possible to utilise *in-situ* mid infrared spectroscopy to determine and monitor a key analyte concentration at various stages during an industrial biotransformation process.

A calibration model was constructed based on five replicates of the biotransformation process, the minimum recommended number of replicates for such a system.⁸³ Since the spectral data substantially outnumbered the available reference measurements the missing values were interpolated based on the experimental determination that the process proceeded via first order kinetics with respect to alanine. This interpolated data along with the first derivative spectral data was utilised to construct a calibration model for the system that returned low errors of calibration and cross validation (19.3 mMol dm⁻³ & 19.4 mMol dm⁻³). Evaluation of the plots of measured concentration against predicted concentration suggested there was no bias within the model to either over or under prediction of the alanine concentration.

External validation of the model was carried out using a replicate of the process that had previously been unseen by the calibration dataset. Predicted concentrations were compared with the interpolated data and an error of prediction determined. The model successfully predicted the alanine concentration of the unseen replicate (RMSEP of 21 mMol dm⁻³).

Efforts were in place to deal with the issue of co-linearity in the system between alanine and pyruvic acid. Break down of the pyruvic acid by hydrogen peroxide to acetic acid however resulted in the degradation of these pyruvic acid spikes. The establishment of an equilibrium where pyruvate concentration remained relatively constant at a low level meant it was difficult to accurately quantify these levels using the developed methods and there was little benefit in the generation of a model for an analyte that didn't exhibit substantial, quantifiable variation over the course of the bioprocess.

7. Fermentation for *In-situ* Biotransformation

All of the biotransformation processes that have been discussed so far have introduced the biocatalyst into the system via the freeze-dried cells. This approach allowed for the long-term storage of the cells until required for the biotransformation process. However the freeze-drying step introduced an energy intensive and time consuming step into the process that was unsuitable when carrying out the process on an industrial scale.

When it was necessary to carry out the biotransformation process on a large or industrial scale it was a more appropriate approach to carry out the fermentation process and immediately utilise the cells after induction, removing the freeze-drying stage of the process.

7.1 Process Summary

In order to carry out the transaminase biotransformation process for the generation of compound B from alanine and compound A it was first necessary to carry out the fermentation process. A fed batch fermentation was carried out with the transaminase enzyme expressed by a modified *E. coli* strain grown under the conditions described and induced with rhamnose (3.3.5.2).

Over the course of the fermentation process, the near infrared spectrum of the fermentation broth was collected at 30-minute intervals. Monitoring key process parameters such as biomass levels, carbon source concentrations and fermentation product concentrations using near infrared have been well documented adopting both an *ex-situ* and *in-situ* approach to sampling.^{18, 23, 24, 27, 32, 35} The ability to offer monitoring and control throughout the bioprocess, from fermentation through to biotransformation, would offer greater process control, process information and quality control in line with the QbD approach of the PAT initiative when compared with the current off-line approaches employed during both these key stages.⁶

7.1.1 Aim & Objective

The aims and objectives of this work was to investigate the potential of using infrared spectroscopy as a mechanism of monitoring the complete bioprocess – from the fermentation stage right through to the biotransformation.

As has been previously discussed the use of infrared spectroscopy to monitor key parameters of submerged culture processes has been extensively reported (1.3.1). In these situations the resulting models have always focused on linking spectral changes with a particular analyte or physical property. However, changes in the spectroscopic response as a result of the system as a whole can often result in a process trajectory or process ‘fingerprint’ for the system. In this case all the physical and chemical changes in the system are modelled, by defining a point where a particular fingerprint is obtained a control strategy could be implemented to improve process efficiency. Such a control strategy is possible because the system is dealt with as a complete unit and the model is not focused on one, or a number, of specific analytes.

To demonstrate that spectroscopy could be utilised to monitor the bioprocess during this fermentation stage, initially models for biomass levels were constructed for the system in question. The errors of calibration and cross validation obtained were compared with values documented in the literature in the first instance.

Using the acquired spectra a control strategy based on the process fingerprint was also to be investigated. Pattern recognition algorithms were considered with a focus on attempting to identify the point when key stages in the process were reached, allowing initiation of subsequent phases resulting in a more streamlined process.

7.1.2 Novelty

Monitoring of submerged culture processes have to date focused on linking spectral variation with a particular analyte or physical parameter. In such a system extensive replicates of the process, and the associated off-line reference analysis, are required

to construct a model. Pattern recognition techniques in conjunction with spectroscopic techniques have been reported as powerful techniques for the classification of raw materials in a variety of industries.¹⁴⁶⁻¹⁴⁸ This work attempts to use these algorithms as a novel method of identifying and controlling the main stages in a submerged culture bioprocess, resulting in a more streamlined and efficient bioprocess that eliminates some of the short lag phases common in such processes.

7.2 Biomass Model

Biomass concentration was one of the key parameters of interest during the fermentation process. With the system currently under investigation the aim of the fermentation process was to generate a sufficient biomass for the production of the required quantities of transaminase enzyme to carry out a large-scale (in this case approximately 10 L) biotransformation process. With successful models for monitoring biomass widely reported, initially an attempt was made to construct similar models for this particular process.^{24, 25, 30, 31}

The acquired near infrared spectra were converted to the second derivative format to enhance the spectral features and reduce baseline drift.²⁴ Samples were removed at various points during the process, to determine if a particular point in the fermentation had been reached, and the dry cell weight determined (Table 23). Immediately following inoculation, the dry cell weight of the system was assumed to be zero since the method used (3.3.5.2.1) could not accurately determine such a low biomass level.

Table 23 - Dry cell weights determined for samples removed during the fermentation process*.

Replicate 1				Replicate 2			
Time (hours)	DCW (g/L)	Std. Dev. (g/L)	Phase	Time (hours)	DCW (g/L)	Std. Dev. (g/L)	Phase
0	0		Inoculation	0	0		Inoculation
14.7	3.4	2.2	Start Feed	13.5	3.5	1.5	
16.7	5.3	1.5		15.65	3.8	1.7	Start Feed
22.0	20.2	1.0	Induction	18.6	7.7	1.9	
39.0	34.3	2.7	Biotransformation	22.1	18.4	0.9	Induction
				39.1	32.7	1.1	Biotransformation
Replicate 3				Replicate 4			
Time (hours)	DCW (g/L)	Std. Dev. (g/L)	Phase	Time (hours)	DCW (g/L)	Std. Dev. (g/L)	Phase
0	0		Inoculation	0	0		Inoculation
15.6	4.3	0.6	Start Feed	15.2	4.2	0.9	Start Feed
18.1	7.2	0.3		17.6	6.6	1.2	
20.1	12.6	1.2		20.1	13.9	0.7	
22.3	19.0	0.6	Induction	22.8	20.7	1.3	Induction
39.4	31.3	1.4	Biotransformation	40.0	32.5	0.5	Biotransformation

*Fermentation processes and biomass measurements generated at University of Strathclyde by Dr. I. Voulgaris.

Previously when an at-line sampling approach had been employed the number of reference measurements was equal to the number of spectra acquired (4,5 & 8). Since the measurements were made *in-situ* in this case the number of spectra acquired outnumbered the reference measurements. To compensate for this the dry cell weights at the time each spectrum was acquired were interpolated, assuming a linear response, based on the limited number of samples available.

A PLS model to monitor the biomass concentrations during the fermentation process was then constructed. The calibration dataset consisted of the first three fermentation replicates with data from the fourth replicate retained to externally validate the resulting model. A total of five latent variables were used in the model, with both the spectral data and reference data being mean centered. Errors of calibration and cross validation, using the leave one out cross validation approach, were 1.7 g/L and 2.1 g/L respectively.

Predicted concentrations for the fourth dataset retained for external validation returned a RMSEP of 2.5 g/L. These values are higher than those previously reported which ranged from 0.2 g/L (Macaloney *et al.* (1994)) to 1.4 g/L (Hall *et al.* (1996) & Arnold *et al.* (2002)).^{24, 25, 30}

During the spectral acquisition and modelling process it was noted that there were some issues with spectral acquisition particularly during the feeding stage of the fermentation. As the biomass increased a number of failed spectral acquisitions were reported. In this situation the spectrometer did not store any spectra and waited until the next spectral acquisition was due.

The reasoning behind these acquisition failures was thought to be a result of the cells clumping together as the biomass levels increased. These clusters of cells temporarily blocked the sample window of the transfectance probe, preventing light from passing through the sample. By the time the next spectrum was due for acquisition the agitation in the system had dislodged the cells. Despite attempts to position the probe in a manner to try and eliminate such issues, these missing spectra

resulted in the calibration model being sparsely populated with spectra representative of the later stages in the fermentation. When the biomass concentration presents a problem with the transmittance probe it may be necessary to alter the sample gap and path length. An alternative solution may be to switch to a reflectance probe design or consider temporarily segmenting the model to improve the predictions during the initial batch phase.^{21, 34}

These results did however suggest that the near infrared spectra could be utilised to monitor and predict the biomass concentration during this fermentation process. Results obtained exhibited slightly higher errors than had been previously noted, however this model was based on a relatively limited dataset and utilised the whole spectral region and not specific regions related only to biomass as had been reported previously.^{24, 25}

7.3 Fingerprinting & Control

7.3.1 Initial Observations

During the attempts to construct a biomass model for the fermentation stage of the bioprocess, it was noted that over the multiple fermentation runs a clear process fingerprint was observed in the spectra.

Carrying out PCA on the spectral data from each of the fermentation replicates and examining a plot of PC1 scores plotted against sample number the key stages of the fermentation could be identified (Figure 7-1).

The initial growth phase was clearly identified, followed by the point where glucose levels within the system became limited. This was indicated by a point of inflection on the PCA scores plot. When feeding commenced the PCA scores increased again before beginning to plateau once the induction stage had begun.

These features identified in the PCA plot were in good correlation with the dissolved oxygen profile obtained during the fermentation (Figure 7-1). Dissolved oxygen levels fell to a low level during the initial growth phase of the process. This was followed by a sharp increase in dissolved oxygen levels after approximately 15 hours, indicating glucose limitation in the system. At this point the increase in dissolved oxygen levels was in good correlation with the point of inflection observed in the PCA scores plot. When the feeding regime commenced the dissolved oxygen levels fell sharply again and remained at a low level through the induction phase.

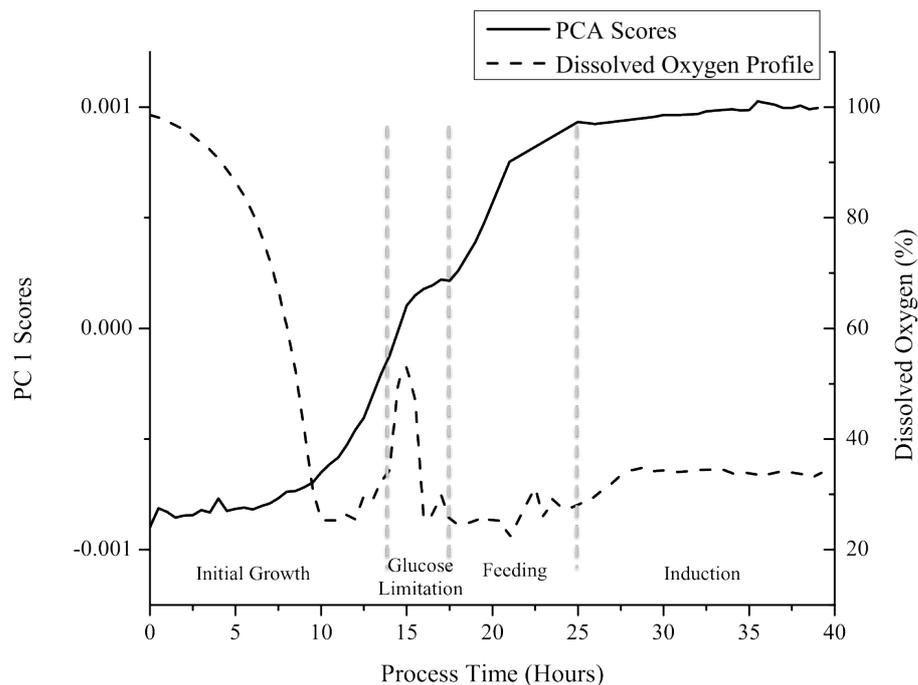


Figure 7-1 - Plot of scores associated with first principal component against process time for third replicate of fermentation process. Clearly identifies the initial growth phase, point of glucose limitation, feeding and finally the point of induction. Also shown is dissolved oxygen profile for fermentation, which confirms the point of glucose limitation by the increase in dissolved oxygen levels.

7.3.2 Pattern Recognition Techniques

Traditionally the modelling approach adopted linked spectral features with a key measured parameter, such as biomass or the concentration of a key media component. Using such models for fermentation control meant a control response was linked to a change in a particular physical parameter. In an example scenario using the biomass model, triggering the start of the feeding regime would be linked to the system reaching a particular biomass level. However in cases that deviate slightly, such as a lower initial carbon source concentration in the batch medium, the trigger point may not be reached as expected and the feeding regime delayed extending the overall time frame for the fermentation and increasing the associated costs.

Infrared spectroscopy, in conjunction with multivariate pattern recognition tools, can be utilised for the analysis and classification of samples. This particular approach is prominent in the food and agriculture industries for the classification of natural products, such as soybean pastes, vinegars and herbs.¹⁴⁶⁻¹⁴⁸ Application of this approach in the pharmaceutical industry for raw material quality control and cleaning validation procedures have also been reported.¹⁴⁹ Examples of the power of such spectral pattern recognition approaches were demonstrated when used in conjunction with surface enhanced raman spectroscopy (SERS), which is a technique complementary to infrared spectroscopy. Using this approach the classification of avian mycoplasmas and influenza viruses have been successfully reported.^{150, 151}

Since a process fingerprint was noted for the fermentation process, a similar pattern recognition approach could be applied for monitoring and controlling the fermentation process. A control system based on this approach would not link the spectra to one particular parameter, and would also offer a time advantage during the model construction phase since the time consuming reference analysis would not be required.

The near infrared spectra acquired during each replicate of the fermentation process were converted to the second derivative format and PCA carried out on the data. The

scores associated with the first principal component were plotted against sample number and each acquired spectrum identified as being representative of the initial growth, glucose limitation, feeding or the induction phases of the fermentation (Table 24).

Table 24 – Classification of each IR spectrum by assigning it to a stage in the fermentation process based on the scores associated with the first principal component.

	Fermentation 1	Fermentation 2	Fermentation 3	Fermentation 4
Initial Growth	1-28	1-25	1-26	1-25
Glucose Limitation	29-30	26-30	27-32	26-30
Feeding	31	31-33	33-36	31-33
Induction	32-56	34-41	37-55	34-42

Having classified the data in this manner a variety of different algorithms were investigated to identify the various stages of the fermentation process from the near infrared spectra.

7.3.3 Standard Isolinear Method of Class Assignment

As a starting point the standard isolinear method of class assignment (SIMCA) was utilised. Each acquired spectrum was assigned a classification based on the identified phase of the fermentation process (Table 24). The constructed SIMCA model was based on the first three replicates of the fermentation process, with the fourth replicate excluded for use as an external validation dataset.

The second derivative spectra were mean centered before the model was constructed, the number of principal components associated with each class determined (Table 25) and used to construct the SIMCA model for the fermentation.

Table 25 - Number of principal components associated with the various stages of the fermentation process used to construct the SIMCA model for the fermentation.

Class	Phase	Number of Principal Components
1	Initial Growth	4
2	Glucose Limitation	3
3	Feeding	3
4	Induction	6

Spectra from the unseen validation dataset (fourth replicate of the fermentation process) were input into the SIMCA model and the process stage for each spectrum identified (Figure 7-2). These predicted values were then compared with the values determined by PCA to assess how well the model was performing and the feasibility of using this approach for process monitoring and control.

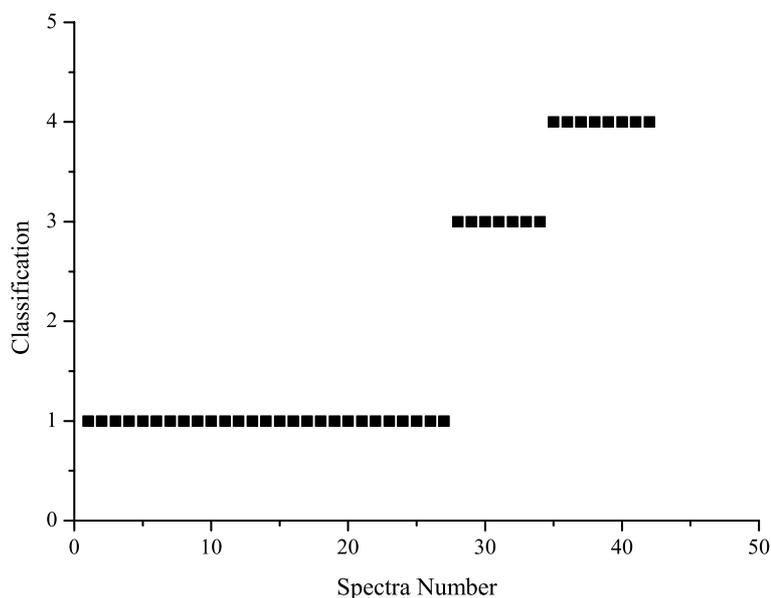


Figure 7-2 - Predicted classifications of the various stages in the fermentation process for the spectra from the fourth replicate of the process designated for validation purposes.

For the fourth replicate of the fermentation process, the SIMCA model predicted that spectra numbers 1 through to 27 represented the initial growth stage, 28 through to 34 represented the feeding phase and 35 through to 42 the induction stage. The SIMCA model did not indicate any of the spectra were indicative of glucose limitation during the process despite the PCA suggesting spectra 26-30 fell into this category.

With the exception of the glucose limitation stages, the classifications of the remaining spectra were in reasonable agreement with those identified from the PCA scores plots. A possible explanation as to why the model failed to correctly distinguish between the glucose limitation and feeding phases may be due to the number of spectra represented in each dataset. Suspected biomass pellet formation particularly during the feeding stage causing interference and blockage of the sample gap resulted in a number of failed spectral acquisitions. Since the calibration model only contained seven examples of spectra in this phase of the fermentation it was possible there were insufficient examples to allow the model to correctly distinguish these spectra.

An alternative algorithm may yield better results for identification of the various stages of the fermentation based on the acquired infrared spectra.

7.3.4 Partial Least Squares – Discriminant Analysis

The partial least squares discriminant analysis (PLS-DA) approach was also considered for monitoring the fermentation process. As before each spectrum was assigned to an appropriate class corresponding to the stage of the fermentation process it represented. All the second derivative spectral data were mean centered. The optimum number of latent variables for the model was determined by an internal validation procedure using the venetian blinds approach (similar to LOOCV except instead of a single sample being temporarily removed it is a block of data), with a total of four latent variables selected for this particular model.

In order to test the constructed model the fourth fermentation replicate, designated for validation and not included in the model, was utilised. The PLS-DA approach assigns each spectrum belonging to a particular class a value of one and those that do not belong to a class a value of zero. In this respect the output from the PLS-DA for the validation dataset was unlikely to be as distinct as a value of either one or zero. Therefore the values were rounded to the nearest whole value, with responses of 0.5 or greater classified as 1, and those below 0.5 a zero (Figure 7-3).

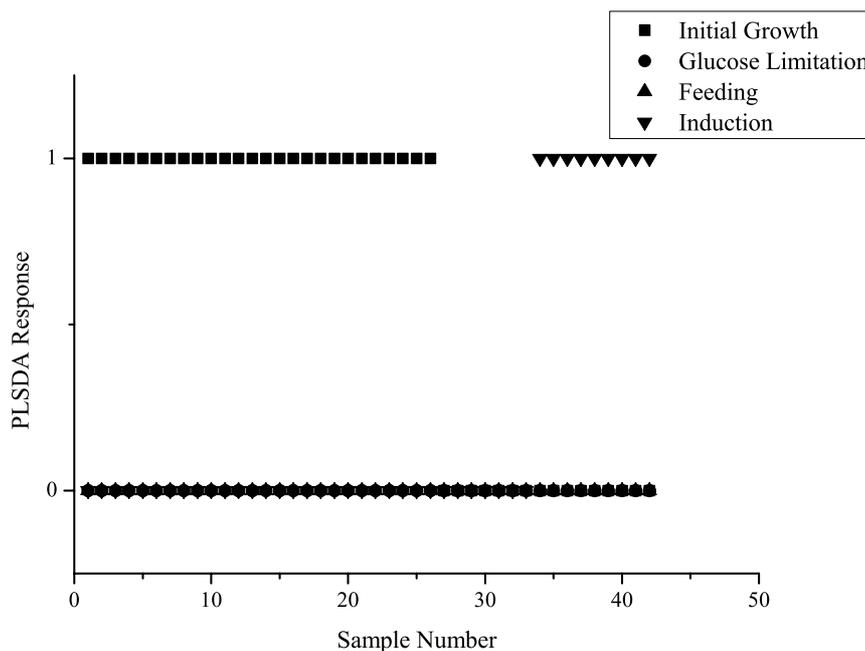


Figure 7-3 – Responses for the external validation dataset predicted using the PLS-DA model constructed for the fermentation stage of the process.

The PLS-DA model suggested that spectra numbers 1 through to 26 corresponded with the initial growth phase of the fermentation, which was in very good agreement with the classifications based on the PCA scores. However the model failed to identify any of the samples as either belonging to the glucose limited or feeding stages of the fermentation. Finally the induction phase was identified as incorporating spectra 35 through to 42, which again was in good agreement with the assigned classifications.

PLS-DA failed to correctly identify samples that represented the glucose limited and feeding phases of the fermentation. This could be attributed to a lack of spectral examples in this region, due to the high number of failed acquisitions, or arising due to errors introduced as a result of the numerical classification system and rounding.

The failure of the model to identify the glucose limitation phase of the process in particular suggested this approach would not be suitable for controlling the fermentation process. Since the premise of control was to avoid glucose limitation there would be little point in attempting to use an algorithm that was unable to identify spectra that fell within this particular region.

7.3.5 Artificial Neural Network

The final approach to pattern recognition investigated was to utilise an artificial neural network (ANN) to identify the various stages of the process. As with the previous algorithms the full region second derivative near infrared spectra were utilised with the various stages of the process identified according to Table 24.

In common with the PLS-DA approach the classification assignments were numerical, with each spectra associated with a particular class assigned a value of 1 and those not associated with this class assigned a value of 0. This resulted in each spectrum being assigned four numerical values depending on the class or stage of the fermentation process that particular spectrum belonged to.

The neural network pattern recognition tool (Matlab R2011b) was used to construct the neural network model for the fermentation. Replicates 1, 2 and 3 were utilised to construct the model, with this calibration dataset further split: 80% of the samples were designated calibration, for training of the model, 15% were designated as validation, to determine the point where training ceased, and the remaining 5% designated as test samples to independently test the constructed model. The test dataset was reduced from the recommended 15% since the samples from a complete

replicate of the fermentation had been retained for this purpose. In doing so this increased the population of the training set by 10% over the default values.

With the ANN model constructed the second derivative spectra from the fermentation designated as validation were input. Predicted outputs were plotted against sample number to identify the various stages of the fermentation (Figure 7-4). An output value of 1 for a particular class suggested this spectrum was indicative of that particular stage in the fermentation process. Due to the error in the model, attaining an output value of either 1 or 0 from the model was unlikely therefore the results were rounded to the nearest whole number. The predicted stages for each spectrum were compared against the values designated by the PCA (Table 24).

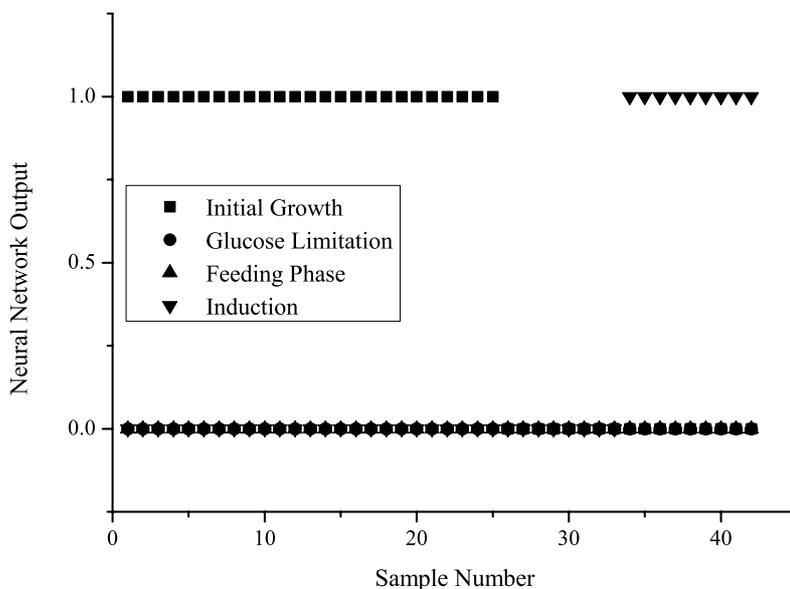


Figure 7-4 - Output from neural network for unseen fermentation replicate used to identify which phase in the fermentation each spectrum relates to.

The output from the neural network was similar to that obtained from the PLS-DA model for this validation dataset. Spectra numbers 1 through to 25 were identified as belonging to the initial growth phase, which was in good agreement with the assigned values. The neural network failed to identify any spectra that represented

the glucose limitation or feeding phases of the fermentation process. Spectra 34 through to 42 were identified as belonging to the induction phase, which was also in good agreement with the assigned classifications.

Arguments already put forward with the other algorithms relating to the lack of spectral data in this region due to the suspected formation of clusters of cells that temporarily block the sample window will again apply in this situation. It was also possible that during the glucose limitation stage in particular there was very little overall change to the matrix as a whole. This lack of change in the system, and therefore by extension the spectra, may be influencing the ability of the network to correctly identify the glucose limitation stage relative to the surrounding stages.

The spectra used in the construction of the above neural network were only the second derivative - no other pre-processing had been applied. With the previous algorithms the spectra were also mean centered. For consistency and in order to draw a comparison with the other algorithms used these second derivative spectra were mean centered and the neural network re-constructed based on this data.

By utilising the mean centered data the neural network model for the fermentation improved substantially. As before the output from the neural network was a numerical value that was rounded to the nearest whole number and used to classify the various spectra (Figure 7-5). Spectra numbers 1 through to 25 were classified as the initial growth phase of the process, which was in agreement with the assigned classifications (Table 24). The mean centered neural network identified that spectra numbers 26 and 28 through 31 were representative of the glucose limitation stage of the process. These values were in good agreement with the assigned classifications where spectra 26 through to 30 represented the cells becoming glucose limited. Spectrum 27 was incorrectly classified as the feeding stage of the fermentation process with no additional spectra identified as feeding in the expected region. Finally the neural network correctly identified spectra 34 through to 42 as representing the induction phase of the fermentation process, which was in agreement with the assigned values (Table 24).

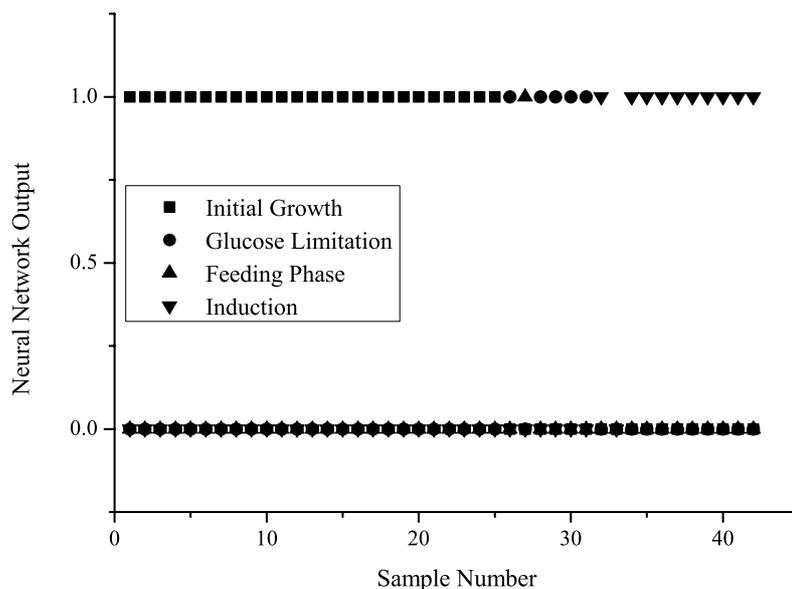


Figure 7-5 - Output for the validation dataset from the neural network constructed using the mean centered second derivative near infrared spectra.

These results have suggested that it would be feasible to utilise an artificial neural network to monitor and control the fermentation stage of this bioprocess. Using the second derivative near infrared spectra it would be possible to identify the point where the cells were beginning to become glucose limited and commence the feeding regime. This would eliminate this point of inflection during the fermentation process resulting in the cells reaching the desired biomass level in a shorter time frame. The approach could also be utilised to identify the point when the fermentation was ready for induction, again potentially streamlining the fermentation process.

7.3.6 Summary

Near infrared spectra acquired during the fermentation process were utilised to predict the various stages of the fermentation. The approach adopted was based on pattern recognition algorithms so that the predictions were not linked with a particular feature or property of the process. Various algorithms were investigated to assess which was the most suitable.

The spectra obtained from three replicates of the fermentation were assigned as representing the various stages of the process and used to construct a model for the system. A fourth replicate of the fermentation was retained for validation of the model. This represents a relatively limited dataset on which to construct a robust model that encompassed the natural variance likely to occur within the system.

Initially a model based on the SIMCA algorithm was considered for the system. With the validation dataset this approach correctly identified the initial growth phase and induction stage of the fermentation. This model failed to identify spectra that were representative of the starving phase incorrectly classifying a number of these spectra as the feeding stage of the fermentation. Since one of the key aims of using this approach was to try and reduce or prevent the cells starving, incorrectly classifying these spectra brought the usefulness of this approach into question.

The PLS-DA algorithm was also considered to try and improve the classifications of the spectra. A model was constructed using the same three replicates of the fermentation and tested using the fourth replicate. In this case the model failed to classify any samples as belonging to either the glucose limited phase or the feeding phase of the process. Since the output from the model was numerical that may be partially attributed to rounding errors, however as with the SIMCA model incorrect identification of glucose limitation in particular suggests this model would not be suitable for monitoring and control of the fermentation.

Use of an artificial neural network with the mean centered second derivative spectra resulted in the best model for the system observed. The neural network correctly identified samples in the initial growth, glucose limitation and induction phases of the fermentation. Samples representative of the feeding phase were not correctly identified however the limited number of samples acquired in this stage due to spectral acquisition difficulties in this region may explain these limitations.

Non-linear algorithms such as ANN's have been applied in industrial settings for process monitoring and control. Lennox *et al.* (2001) summarised a number of

industrial applications of the technique, as well as a number of common issues associated with implementation of the technique.¹⁵² Neural networks are also commonly employed within the food industry. Jiang *et al.* (2011) reported their successful application for the identification of Chinese soybean pastes. Of the three varieties investigated PCA was unable to distinguish between two similar varieties. The ANN however was able to successfully distinguish between all three, with success rates in region of 97% for the validation dataset.¹⁴⁷

Despite the ANN's apparent success with the current system, the 'black box' nature of the mathematical algorithms still casts doubts over the reliability of the algorithm.⁸³

7.4 Conclusions

Using the full region near infrared spectra obtained from the *in-situ* transreflectance probe during the fermentation it was possible to construct a biomass model for this system. Errors of calibration cross validation and prediction for this model were higher than previously reported.^{24, 25, 30} However this can be explained the relatively small number of replicates populating the calibration model and the unrefined spectral regions used to construct the model. Based on these observations it was concluded that replicate models for those analytes described in the literature could be constructed for this particular fermentation process if required.

An alternative approach that would see the time taken to construct a model for the system reduced and allow for a more generalised approach to *in-situ* fermentation monitoring based on spectral pattern recognition was investigated.

Over the course of the fermentation a particular fingerprint was observed that could be utilised to classify each spectrum as belonging to a particular stage in the process. Utilisation of this pattern represented a method of monitoring and controlling the fermentation that was not linked with a specific process parameter, but treated the process as a whole.

Various algorithms for pattern recognition were investigated: SIMCA, PLS-DA and artificial neural networks. The best results were obtained using a neural network for pattern recognition. In this case the system was able to correctly identify the point at which the cells became glucose limited, a stage the SIMCA and PLS-DA approaches were unable to correctly identify. Applied in real time, an indication that this point had been reached would then allow for corrective action, in this case feeding, to be taken eliminating the starving phase and streamlining the process. Identification of when the system was ready for induction was also possible and had been correctly predicted using all the pattern recognition algorithms employed.

Whilst the constructed model does require some additional development, it does indicate the feasibility of monitoring and controlling the fermentation process using a neural network for pattern recognition.

8. *In-situ* Biotransformation Process

8.1 Process Overview

The particular biotransformation process studied was a transaminase process for the generation of the chiral amino acid product compound B. A transaminase enzyme acts by transferring the amino functional group from a donor molecule, in this case L-alanine (absolute stereochemistry is S-(+)-alanine) to replace the keto functional group of the acceptor molecule compound A (Figure 8-1). The process occurs with the retention of stereochemistry, so the S-configuration of the alanine was transferred to compound A forming the S configuration of the compound B product.

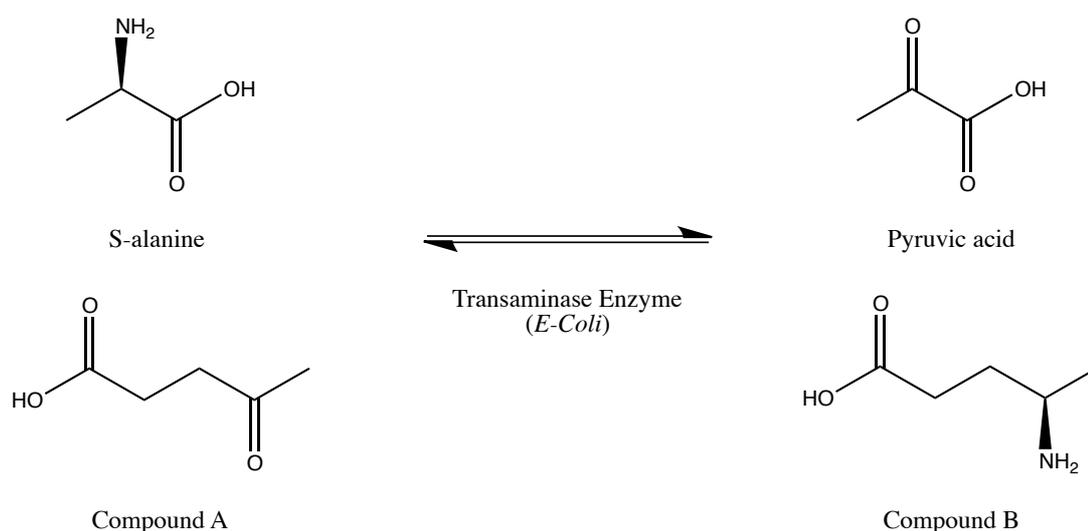


Figure 8-1 - Summary of the biotransformation process for the generation of compound B from S-alanine/L-alanine and compound A substrates proceeding via transaminase enzymes.

The biotransformation process was carried out immediately following the induction phase of the fermentation process. All process parameters, such as agitation and aeration, and controlled parameters, such as pH, were maintained at the conditions utilised during the fermentation stage of the process (3.3.5.2). Full details of the

experimental procedure employed for this biotransformation process are detailed in Section (3.3.5.3).

Generation of compound B is in equilibrium, so to obtain the maximum yield from the biotransformation process the equilibrium position needs to be driven in favour of the products to make the process viable.

During the fermentation stage the organism (a modified *E. coli*) was supplied with a monosaccharide carbon source via the batch media and also in a more concentrated form via the glucose feed. This monosaccharide was utilised by the cells for the generation of energy via the glycolysis metabolic pathway (Figure 8-2).¹⁵

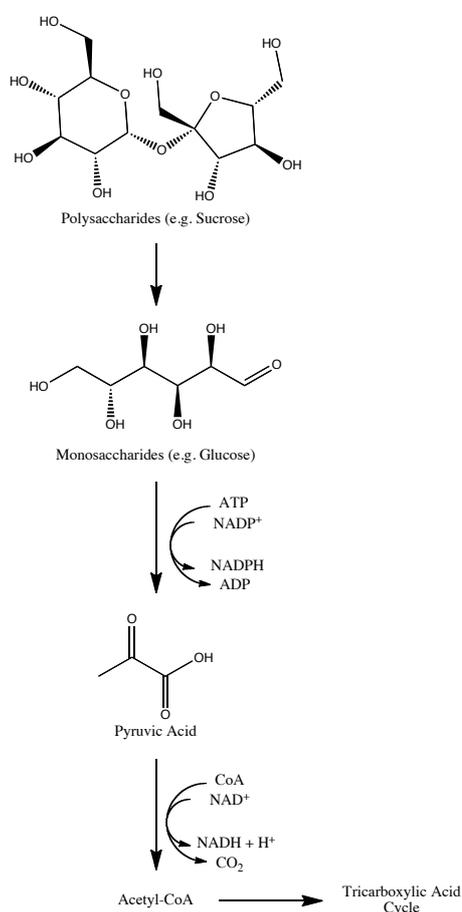


Figure 8-2 - Pathway illustrating the breakdown of polysaccharides to a monosaccharide and then to pyruvic acid and onto the TCA cycle adapted from Prescott *et al.*¹⁵

When the biotransformation process was underway the supply of the monosaccharide glucose via the feed had ceased. In order to keep the cells alive a carbon source was required to continue to allow the generation of energy by this metabolic pathway. As had been previously noted (Figure 8-1) pyruvic acid was generated as a by-product of the biotransformation process, which could then be utilised by the organism for the generation of energy.¹⁵ This utilisation of the generated pyruvic acid also resulted in a shift to the equilibrium position. In order to compensate the system adjusts to replace this pyruvic acid, driving the equilibrium in favour of the desired product, in this case compound B.

8.2 Aim & Objective

As with the various other biotransformation processes that have been investigated (Sections 4, 5 & 6) process measurements have been relatively limited, particularly near real-time measurements. Many of the key process parameters specifically the analyte concentrations were determined using off-line methods.

Like the other processes that have been investigated the objective of this work was to evaluate the potential of near and mid infrared spectroscopic techniques as a means of rapidly determining key analyte concentrations as the biotransformation process progressed.

Near infrared spectroscopic data was obtained *in-situ* making use of a transreflectance probe inserted directly into the bioreactor system (1.2.1.5 & 3.5.2). Mid infrared spectroscopy was investigated adopting an at-line sampling approach (3.4.1). Depending on the feasibility of these results, the approach could also be applied *in-situ* should the equipment be available.

8.2.1 Novelty

Novel aspects of this work relate to the biotransformation stages position within a larger bioprocess. Successful construction of quantitative models for the

biotransformation process would result in infrared spectroscopy being utilised to monitor and control the entire bioprocess, from the fermentation stages through to the biotransformation.

This work also represents the novel use of the LSS calibration transfer algorithm not only to a complex sample matrix but also to a biological sample matrix. Previously published applications of the algorithm relate to fairly simple sample matrices such as pharmaceutical tablets or solvent mixtures.¹²³ The complex nature of the biotransformation matrix represents a substantial challenge over the currently reported applications.

8.3 Analytical Reference Method

Prior to any replicates of the biotransformation process being carried out, it was essential that robust and reliable analytical reference methods were in place to accurately quantify the key analytes of interest in the biotransformation process.

8.3.1 Preparation of Standards

The compound B product of the transaminase biotransformation process was not commercially available as a reference standard for the quantification of this component in the biotransformation samples. As such the material was prepared and purified based on an adaptation of the approach outlined by Clark *et al.* (2001) (3.1.2.3).¹²⁵

The desired product was purified using a flash chromatography column and eluted using ammonia hydroxide solution. This fraction was collected and the ammonia hydroxide removed by rotatory evaporation to obtain the crystalline material. Approximately 10mg mass of this material was dissolved in deuterium oxide (Sigma Aldrich) and a proton NMR spectrum of the material collected (Figure 8-3).

Each of the signals in the acquired NMR spectrum and the associated coupling were identified (**Table 26**). The observed chemical shift values and multiplicity of the signals due to the main functional groups were consistent with the theoretical values for compound B.

Based on this proton NMR, it was concluded that the desired product had been generated by the procedure adopted. The absence of any additional signals in the NMR spectrum also suggested that the flash chromatography procedure employed had resulted in a sufficiently pure sample for use as an analytical standard for the identification and quantification of compound B using the reference analysis method.

The NMR suggests that both the R and S configurations of the product had been generated by this procedure, evident by the splitting of the signals from the protons adjacent to the chiral centre. Although the biotransformation process only generated a single enantiomer, the reference method was unable to distinguish between configurations so the racemic standard will still allow for quantification of the analyte.

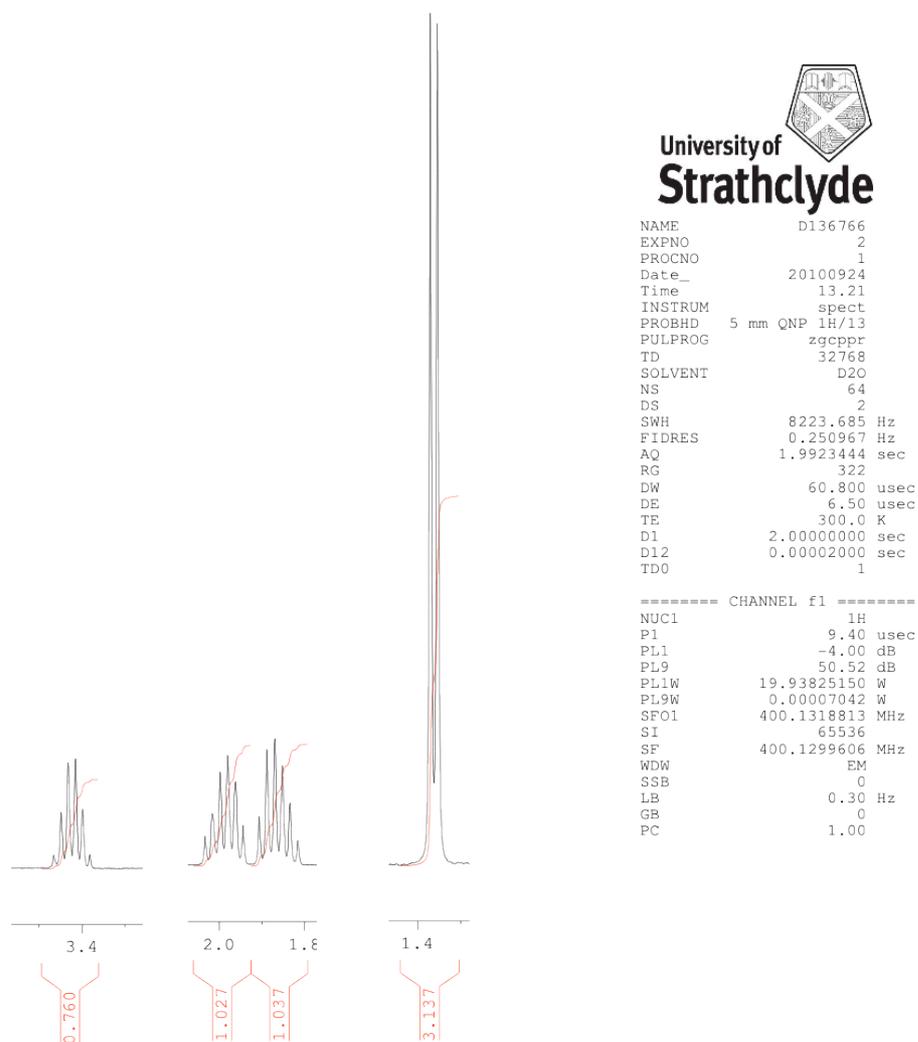


Figure 8-3 – Selected features from NMR spectrum of compound B produced using the adapted procedure outlined by Clark *et al.* (2001) and purified by flash chromatography with a strong cation exchange resin.

Table 26 - Assignment of the selected signals in the acquired proton NMR of synthesised compound B for use as analytical reference material.

Chemical Shift	Multiplicity	Corresponding Proton
1.34-1.36ppm	Doublet	$\underline{\mathbf{R}}_5\text{CHNH}_2\text{R}_4\text{COOH}$
1.81-1.90ppm	Multiplet	$\text{R}_5\text{CHNH}_2\underline{\mathbf{R}}_4\text{COOH}$
1.94-1.99ppm	Multiplet	$\text{R}_5\text{CHNH}_2\underline{\mathbf{R}}_4\text{COOH}$
3.39-3.44ppm	Sextet	$\text{R}_5\underline{\mathbf{C}}\text{HNH}_2\text{R}_4\text{COOH}$

The synthetic procedure returned a relatively poor yield of racemic compound B product with a crude yield of approximately 9% based on the mass of substrate utilised assuming 100% conversion of reactant. Despite this poor yield the process generated sufficient crude and purified compound B product to allow the analysis to be carried out.

8.3.2 Amino Acid Quantification

In this particular case it was desirable to quantify both the amino acids of interest (L-alanine and compound B) using a single HPLC method. As had been the case previously both these compounds required a pre-column derivatisation stage to be employed to allow detection.

The gradient elution method employed in this case was that detailed for the separation and quantification of all discussed amino acids with the exception of TBG (3.2.1.2.2). Under these conditions sufficient baseline separation between the L-alanine and compound B was achieved with both analytes also exhibiting good, symmetrical peak shapes. Alanine returned a peak asymmetry factor of 0.94 and a tailing factor of 0.97 whilst the compound B product returned an asymmetry factor of 0.92 and tailing factor of 1 (Appendix II).

As there were two amino acids of interest in these samples it was necessary that the derivatisation reagents were present in sufficient excess to ensure reaction with all

the amino acid content present in the sample. Presence of the reagent in excess ensured that the derivatisation process was consistent and that a competing reaction between the amino acids was not occurring. Based on the highest total concentration of amino acids in the calibration standards ($\sim 19 \text{ mMol dm}^{-3}$), even at the highest concentration the derivatisation reagent was present in excess relative to the amino acids ($\sim 30 \text{ mMol dm}^{-3}$).

8.3.2.1 Method Validation

Validation of the method was subsequently carried out to assess the error associated with quantification using the developed method. A series of calibration standards covering a concentration range from 2 mMol dm^{-3} through to 9 mMol dm^{-3} for both the amino acids were prepared and analysed by the method. Ten replicate injections of a sample containing a mixture of L-alanine and compound B were also injected to assess the reproducibility of the developed method and estimate an error associated with quantification (Equation 16, Equation 19 & Equation 20)

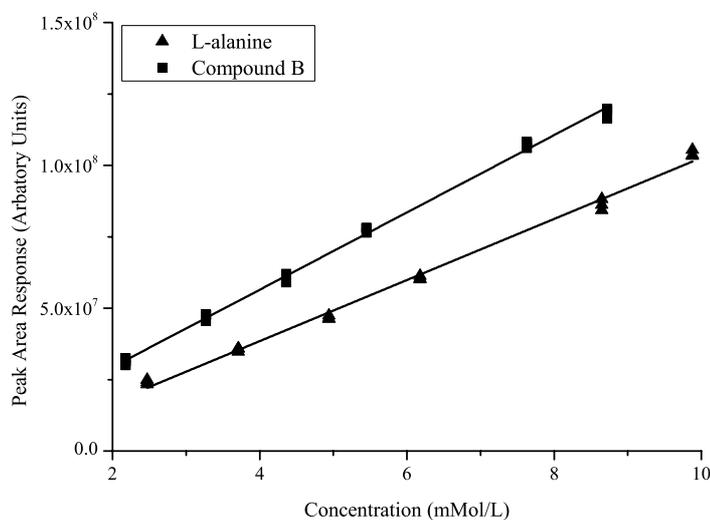


Figure 8-4 - Calibration curves for the quantification of L-alanine and compound B using the developed HPLC method.

Based on these results the error associated with the developed method was determined as being $\pm 0.20 \text{ mMol dm}^{-3}$ for L-alanine and $\pm 0.18 \text{ mMol dm}^{-3}$ for compound B. (Appendix II).

8.3.3 Organic Acid Quantification

Quantification of compound A was achieved using the HPLC method described with the Rezex column and refractive index detector as described (3.2.2.2). Conditions for this method were relatively simple, an isocratic mobile phase composition of 0.005N H_2SO_4 at a flow rate of 0.5 mL/min, as indicated by the column manufacturer. It was decided at this point that only compound A would be quantified by the method as it was unlikely that significant quantities of pyruvic acid would be observed due to its metabolism by the cells.

8.3.3.1 Method Validation

To confirm that a linear response was obtained, a series of calibration standards at varying concentrations between 5 mMol dm^{-3} and 65 mMol dm^{-3} were prepared and analysed. A linear line of best fit was plotted through the data points and the equation of the line and correlation coefficient determined ($r=0.999$) (Figure 8-5).

As with the other methods, ten replicate injections of a sample were carried out to assess the reproducibility of the method. Based on these results, and the calibration data, an estimation of the total error associated with the method for quantifying compound A was determined as being $\pm 0.70 \text{ mMol dm}^{-3}$ (Appendix II).

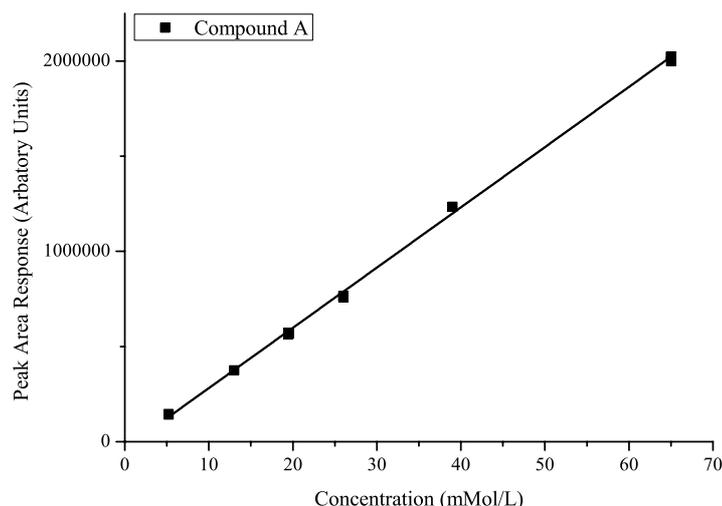


Figure 8-5 - Calibration curve for the quantification of compound A using the described HPLC method.

8.4 Initial Feasibility

To determine if monitoring the biotransformation process was feasible using *in-situ* near infrared or at-line mid infrared spectroscopy the pure component spectra of the three key compounds were acquired and examined.

8.4.1 In-situ Near Infrared

Separate standards of L-alanine, compound A and compound B were acquired using the Foss XDS spectrometer system (3.5.2) and attached transmittance probe with a 0.5 mm window gap giving an effective path-length of 1 mm. The acquired spectra were exported and converted to the second derivative spectrum before being examined to identify potential regions of interest. Due to the absence of the combination bands region, since there is absorption of this frequency region by the silica fibre optic bundle¹⁴¹, the main area of focus was the first overtones region between 5000 cm⁻¹ and 6660 cm⁻¹ (Figure 8-6).

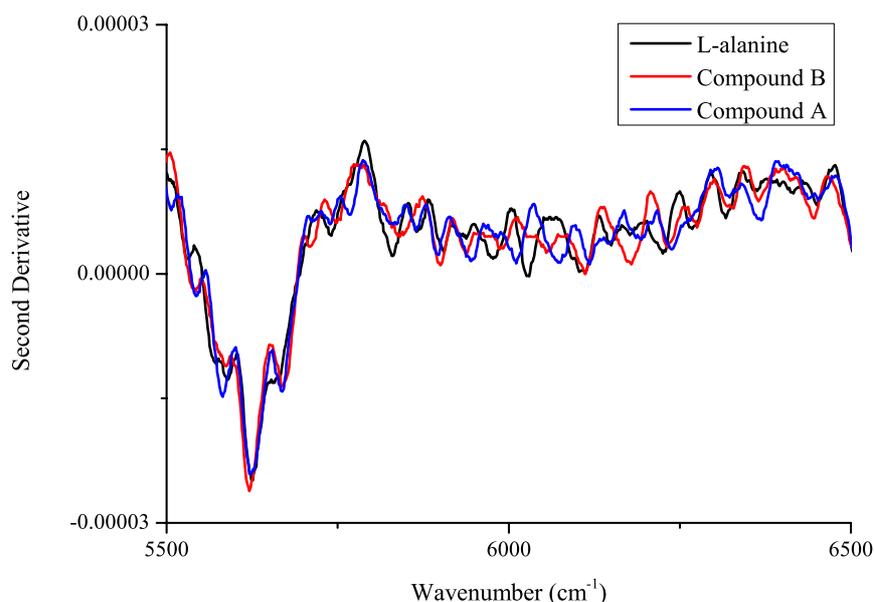


Figure 8-6 - Second derivative pure component spectrum of key analytes of interest in the first overtones region.

Pure component spectra of pyruvic acid were not acquired since this component was not expected to be present at substantial concentrations due to metabolism by the cells.

Within this considered region it was evident that there were a number of close similarities between the acquired infrared spectra of these compounds, particularly between compounds A and B. This is particularly obvious in the approximate regions below 6000 cm^{-1} and above 6250 cm^{-1} . Differences in the acquired spectra were also noted, with unique features due to L-alanine observed just above 6000 cm^{-1} and a unique feature of compound B noted at approximately 6200 cm^{-1} . However, given the previous issues noted with applying the near infrared region to these biotransformation processes it was questionable whether the near infrared region could be successfully utilised to construct an independent quantitative model for the process.

These feasibility samples were presented in an optically clear matrix, however the actual process samples immediately followed the fermentation process and so would contain a high content of cellular material. Cellular material may have an interfering effect on the spectra due to an increased light scattering as the cell density increases.³⁰ To counter that argument, the cellular density would already have reached its maximum at the point when the biotransformation process began. In addition the selected probe design, which was essentially a hybrid of transmission and reflectance probe designs, should still allow for spectral measurements to be made even at high biomass levels.¹⁷

Since this represented the *in-situ* application of the spectroscopic technique, an additional factor that required consideration was the increased noise in the acquired spectra resulting from the required agitation and aeration of the system. Given the weak, and non-specific nature of the technique any increase in spectral noise may obscure the already weak spectral features in the identified regions of interest.

In an attempt to determine if these parameters had an influence on the quality of the acquired spectra, the near infrared spectrum of water in the first overtones region of interest was collected at various agitation rates. Spectra were converted to the second derivative format and the effect on the baseline noise evaluated (Figure 8-7). Water was selected because it was the major component of the fermentation media and was present in large quantities during the biotransformation. In addition the OH vibrations of water molecules tend to dominate in the near infrared region, so the effects of agitation and aeration on these strong spectral features would give a good indication as to how the weaker spectral features would be affected.

These results suggest that, even at relatively low agitation speeds, background noise introduced as a result of mechanical vibration within the system has begun to have an effect on the quality of the acquired spectra.

Based on these observations, with the agitation rate and the limited distinctive spectral features for each of the key analytes, it was concluded once again that it was

unlikely the near infrared region could offer sufficient resolution to allow the construction of a robust multivariate model that could independently predict the concentrations of the three key analytes of interest during the biotransformation process. Based on previous experience it was felt that constructing a model based on the mid infrared spectroscopic region would be more successful for this particular biotransformation system under investigation.

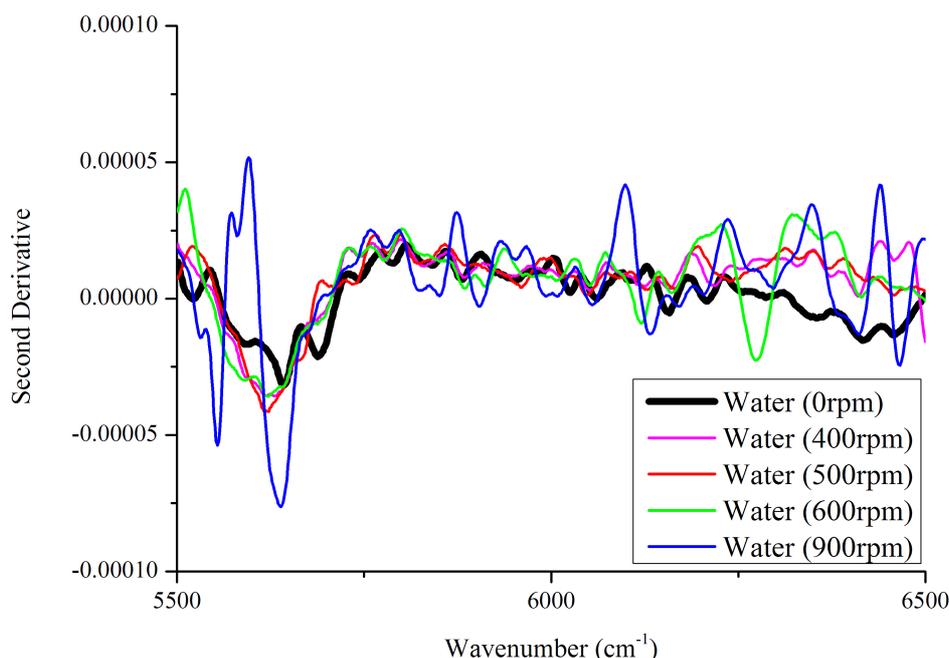


Figure 8-7 - Near infrared spectrum of water acquired using *in-situ* transmittance probe at a variety of agitation rates likely to be experienced during the biotransformation process to assess the effect on background noise in the spectra.

8.4.2 At-line Middle Infrared

Moving into the mid infrared region not only were the specificity issues addressed by the inclusion of the fingerprint region, but many of the problems noted due to increased noise due to agitation and aeration, as well as the increased biomass levels, should be reduced. This was due to the substantially smaller path-lengths attained with the utilisation of the ATR crystals. With reported effective path lengths in the

region of 10 μm attained from multi-bounce ATR crystals, and typical bacteria cells being in the region of 1.3 μm – 4 μm in size, interference effects were unlikely to present the same issues with this sampling mechanism.^{15, 53} Although the mid infrared system utilised was an at-line system where reactor environment would not be a concern, the arguments put forward will still apply should an *in-situ* mid infrared probe be employed.

Each of the pure component samples of the three key analytes of interest had their spectra acquired in the mid infrared region as described (3.4.1). The acquired spectra were converted to the first derivative to enhance spectral features and reduce baseline drift.

As had been the case previously many of the redundant spectral features were removed focusing mainly on the fingerprint region and locations where the amine signals were expected (950 cm^{-1} through to 1760 cm^{-1}) (Figure 8-8).

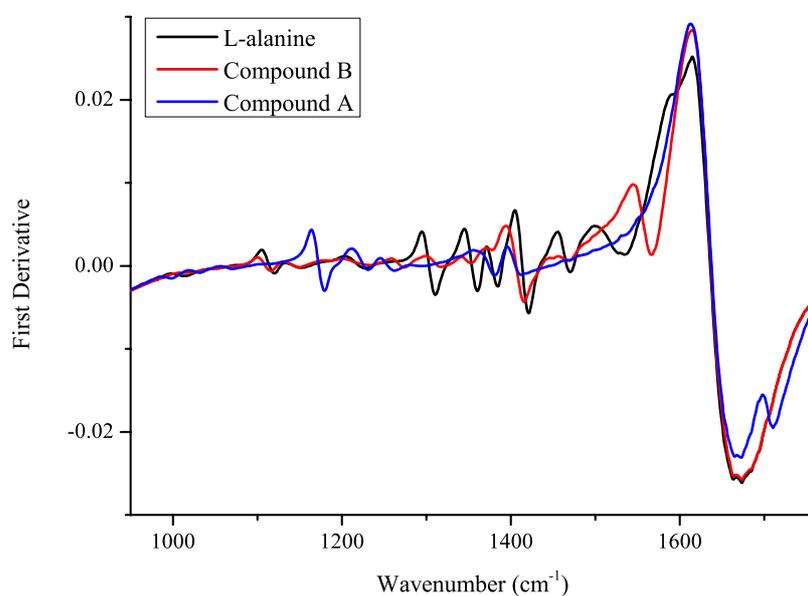


Figure 8-8 - First derivative spectrum of pure component samples of key analytes in the biotransformation process. Main region of interest shown between 950 cm^{-1} and 1760 cm^{-1} .

Overlapping spectral features were noted for both the amino acids of interest, with subtle shifts between the two observed in the region between 1300 cm^{-1} and 1500 cm^{-1} . Points of difference arising from the amine functional groups were also noted between 1500 cm^{-1} and 1600 cm^{-1} with each amino acid giving rise to a unique signal in this region. Many of the signals observed from compound A were noted between 1200 cm^{-1} and 1300 cm^{-1} and did not appear to overlap with any signals from the amino acid.

Based on these observations, it appeared there were sufficient variations in the acquired spectra of the three key analytes to allow the fingerprint region of the mid infrared to be used to construct a quantitative multivariate model for the system.

8.5 The Biotransformation Process

A total of four replicates of the fermentation and *in-situ* biotransformation process were carried out according to the procedure described (3.3.5.3). Samples removed at regular intervals during the process were heated at 100 °C for a three minute period to denature the enzyme and prevent further conversion. These samples were then stored under refrigerated conditions until the mid infrared spectra were acquired.

For each sample the mid infrared spectrum was acquired in triplicate and each sample was also quantified using the corresponding HPLC method, again in triplicate. Making triplicate scans of each sample increased the overall number of spectra in the calibration model but also served as a means of equipping the model to deal with the subtle variations that may occur between spectra on making replicate measurements.

8.5.1 Middle Infrared Model

Ruckebusch *et al.* (2002) considered five replicates of a similar biological process as an appropriate number of replicates to ensure the natural variations observed between process replicates were adequately represented in the model.⁸³ Given that only four replicates were available for both calibration and validation purposes it was

decided that the data would be split to give three process replicates for calibration with the fourth retained for external validation of the constructed model.

In order to determine which of the limited replicates of the process were to be utilised for calibration and validation, PCA was carried out on the spectra and used to determine the most appropriate datasets ensuring that the calibration dataset encompassed as much of the variability in the system as possible. This was based on the assumption that natural variations in the process were represented in the acquired spectra.

PCA was carried out on the first derivative fingerprint region (950 cm^{-1} to 1760 cm^{-1}) spectra from the four process replicates. The first derivative data was mean centered, however no additional pre-processing or manipulations were carried out.

A plot of scores associated with the first latent variable against the scores associated with the second latent variable was constructed (Figure 8-9).

The spectra obtained from the third replicate of the biotransformation process were substantially different with respect to the variance captured by the second latent variable compared with the other replicates. Since this was a deviation from what had been observed with the other process replicates, a visual inspection of the spectra was carried out to confirm that this represented a true variance in the process and was not due to a spectral anomaly. There did not appear to be any irregularities in the acquired spectra for this process replicate compared with the other replicates therefore suggesting this did represent a true variation in the system and should be incorporated in the model.

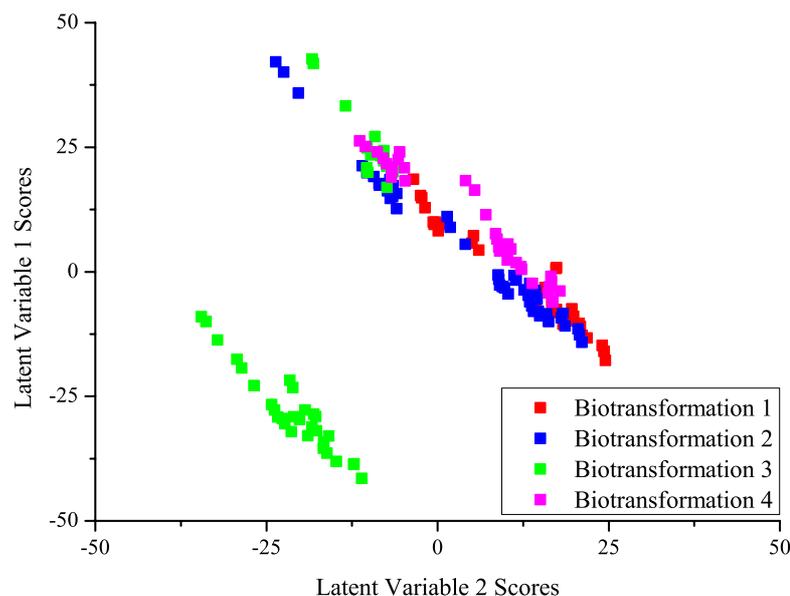


Figure 8-9 - Latent variable 1 scores plotted against latent variable 2 scores to identify the process replicates used in the calibration model thus ensuring the maximum variance observed was included in the calibration model.

Based on the PCA, biotransformation replicates 1, 2 and 3 were selected for the calibration dataset, leaving the fourth replicate for use as an external validation dataset. Selecting the calibration and validation datasets in this manner also ensured that the validation dataset was within the predictive ability of the constructed model given the limited calibration data available to span the potential variance in the process.

Using this first derivative fingerprint region spectra (950 cm^{-1} through 1760 cm^{-1}) of the identified process replicates a PLS model for the system was constructed. The only other pre-processing carried out was to mean centre both the spectral data and the reference data. Leave one out cross validation was also employed as a means of internal validation of the model and to assist in selecting the optimum number of latent variables for the model.

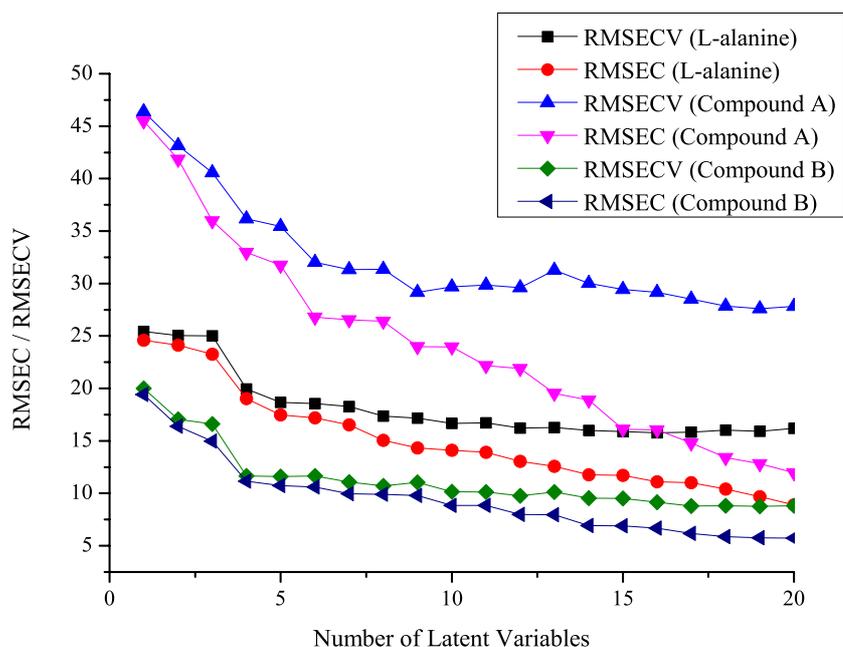


Figure 8-10 - Plot of RMSEC and RMSECV values against latent variable number to determine the optimum number of latent variables for use in the constructed PLS model.

Examination of the plot of root mean square error of calibration/cross validation versus the number of latent variables was used to determine the optimum number of latent variables for the model (Figure 8-10). From this figure, four latent variables appeared to be the most appropriate selection. Arguably there was some suggestion that up to six latent variables may be appropriate, however when increased above four latent variables a clear over-fitting of the data, where more terms than necessary are included in the model¹⁵³, particularly with regards to the compound A concentrations became apparent. Based on these observations the calibration model was constructed with a total of four latent variables.

Errors of calibration and cross validation for the constructed PLS model based on 4 latent variables are detailed in Table 27. There was good agreement between the calibration error and cross validation error for the three key analytes of interest. Examination of a plot of predicted concentration (from cross validation process) against measured concentration (from offline reference method) did illustrate a few

questionable points that were potential outliers, however the vast majority of the points appeared randomly scattered around the central diagonal (Figure 8-11 (i, ii & iii)).

The spectral data from the process replicate designated as a validation dataset was then presented to the constructed model and the concentrations of the three key analytes predicted. Predicted concentrations were then compared with the values obtained using the traditional off-line reference analysis method to calculate a RMSEP value.

Table 27 - Errors of calibration and cross validation associated with the calibration model along with the errors of prediction for the unseen validation dataset.

	RMSEC (mMol dm⁻³)	RMSECV (mMol dm⁻³)	RMSEP (mMol dm⁻³) [r²]
L-alanine	19.0	20.0	57.0 [0.57]
Compound A	33.0	36.2	19.6 [0.71]
Compound B	11.2	11.7	37.7 [0.83]

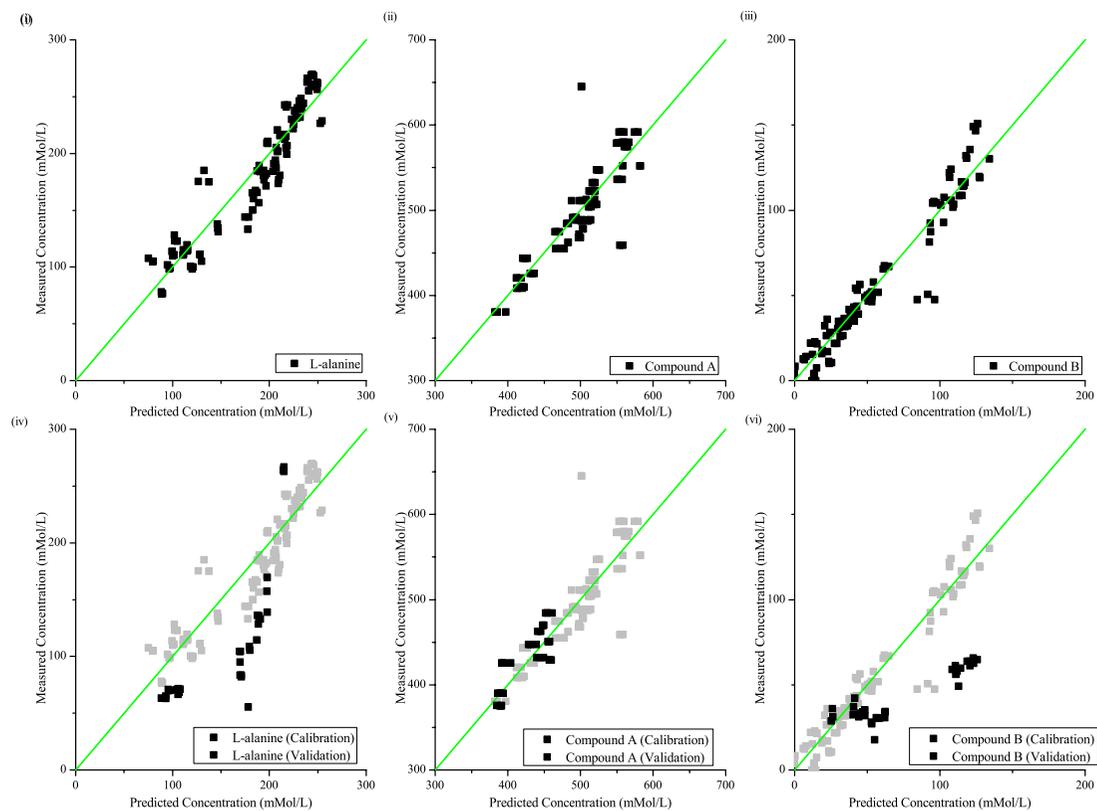


Figure 8-11 - Plots of predicted concentration (mMol dm^{-3}) (from PLS model) against measured concentration (mMol dm^{-3}) (from reference analysis) for the various key analytes of interest for both the internal (i, ii & iii) and external (iv, v, vi) validation procedures.

The errors of prediction associated with the unseen validation replicate were substantially different from the RMSEC and RMSECV values. Examination of the plot of predicted concentration against measured concentration confirmed that the model was largely over-predicting the concentration of L-alanine and compound B in the dataset (Figure 8-11 (iv, v, vi)).

It was expected that the fermentation process immediately preceding the biotransformation process, as well as the slight variations in analyte concentrations added in each case, would provide sufficient variation to prevent co-linearity. The high errors of prediction attained would however suggest that co-linearity had some impact on the predictive ability of the model.

To confirm whether co-linearity was having an influence on the constructed model the associated regression coefficients were examined (Figure 8-12). From this plot it was evident there was co-linearity within the system, both the first and second regression coefficients for the model were identical with the third regression coefficient being the exact opposite of the first two. This would suggest that spectral features arising due to the alanine are being utilised to estimate the concentration of compounds A and B, with the same holding true for the other analytes. Co-linearity within the system would explain the failure of the model to accurately predict the key analyte concentrations for the validation dataset.

8.5.2 Model Refinement

In order to break this co-linearity that existed within the calibration model the samples were spiked by the addition of known concentrations of the pure component materials. Previously the retrospective adulteration of samples was not favoured since this would serve to change the composition of the sample media. In the current situation however an adaptive calibration approach was determined the most appropriate, as the processes had already been run and the samples taken.

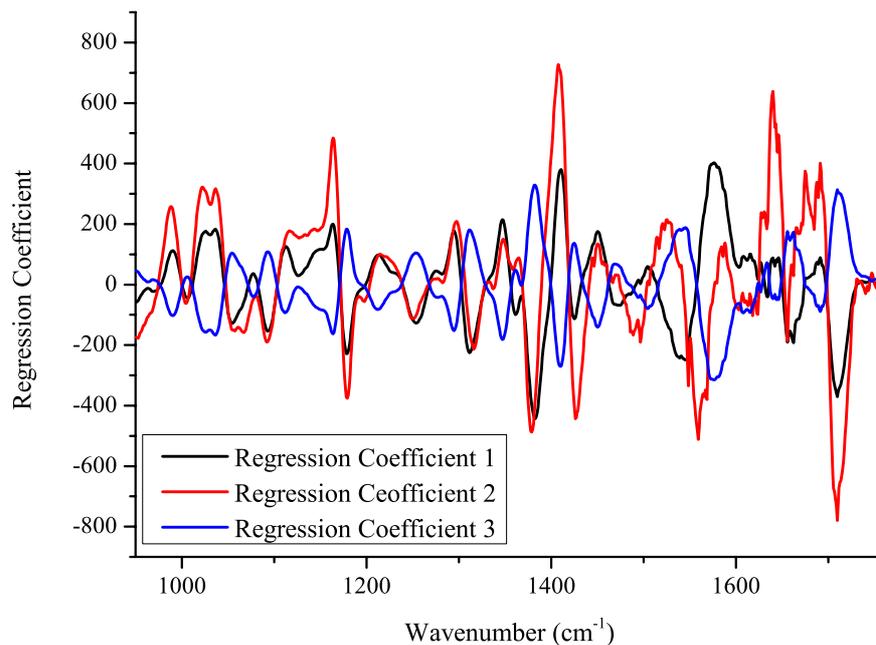


Figure 8-12 - Regression coefficients associated with the constructed PLS model for the biotransformation process.

Spiking of the samples was based on the adaptive calibration approach utilised by Riley *et al.* (1998) with some amendments.³⁸ Rather than the addition of “random and known amounts” of the key analytes, the approach taken was to construct a partial factorial experimental design (three factor, two level D-optimal design) using the Design Expert software.³⁸ The spiked analyte concentrations of L-alanine and compound A varied between 0 and 250 mMol dm⁻³, whereas the concentrations of compound B varied between 0 and 100 mMol dm⁻³ (Appendix III). Lower concentrations of compound B were utilised due to the small quantities of purified product obtained from the synthetic procedure employed to generate the reference standard.

Adopting the approach where samples were spiked according to the generated experimental design ensured that the stoichiometric linkages were broken by design and not random chance or co-incidence.

Spiking of the samples was achieved by introducing small volumes of concentrated solutions of the three analytes of interest and then diluting these samples to a known volume with the actual process sample. In the case of the fifth spike in the DoE, where 100 mMol dm^{-3} of each analyte was introduced, $400 \mu\text{L}$ of each stock solution (concentration $\sim 1250 \text{ mMol dm}^{-3}$) was added and made up to a total volume of 5 mL using the actual process sample. A total of 42 spikes were prepared and encompassed samples from all four replicates of the biotransformation process carried out.

All spiked samples had the analyte concentrations re-quantified using the appropriate reference methods and the spectra of these new samples were acquired. The actual biotransformation samples underwent a dilution, in the region of 1:5. However the spiking process meant that the analyte concentrations covered by the model still encapsulated those expected during the process. In addition the specificity offered by the mid infrared region meant factors not being monitored, such as biomass and other analytes, would not be significantly affected.

Using only the spiked samples and the corresponding reference data the PLS model for the biotransformation process was reconstructed, and the regression coefficients examined to determine whether the co-linearity within the model had been addressed. The PLS model constructed from the spiked dataset returned three unique regression coefficients suggesting that the co-linearity issues previously observed within the original model had been addressed.

Using this PLS model based on the spiked samples, the concentrations of the key analytes in the original replicate designated for external validation were estimated. Given that some of the samples in the calibration model were actually spiked samples from this validation dataset, it was expected that this model would predict the analyte concentrations extremely well, as to some degree these were already samples included in the calibration dataset. The RMSEC, RMSECV and RMSEP values for the model are detailed in Table 28.

Table 28 - Errors associated with the calibration model constructed using the spiked datasets and prediction of the key analyte concentrations in the external validation dataset.

	RMSEC (mMol dm ⁻³)	RMSECV (mMol dm ⁻³)	RMSEP (mMol dm ⁻³) [r ²]
L-alanine	21.5	23.5	174.1 [0.63]
Compound A	58.9	64.0	599.1 [0.87]
Compound B	24.9	27.5	141.3 [0.80]

Errors of calibration and cross validation were slightly higher than had been observed previously (Table 27) however this was unsurprising given that the co-linearity within the model had been addressed. However the errors of prediction associated with the validation dataset were significantly higher than had been observed with the co-linear model.

This calibration model was however based purely on the retrospectively spiked samples. Whilst suitable models for various systems have been constructed based on spiked samples there are reported cases where the introduction of synthetic samples have actually hindered the model.^{17, 38} The lack of actual process samples in the model may offer some explanation for the large errors of prediction observed.

In an attempt to improve the errors of prediction a combined model based on the spiked samples and the original spectra was constructed. The calibration dataset incorporated all the spiked samples along with the samples from the three biotransformation process replicates identified by PCA.

Table 29 - Errors of calibration, cross validation and prediction associated with model constructed from both the spiked samples and the original samples obtained from the biotransformation process.

	RMSEC (mMol dm ⁻³)	RMSECV (mMol dm ⁻³)	RMSEP (mMol dm ⁻³) [r ²]
L-alanine	43.1	45.2	70.4 [0.58]
Compound A	68.2	71.2	64.1 [0.16]
Compound B	36.9	37.9	44.8 [0.78]

Including the original spectra in the model has improved the errors of prediction substantially (Table 29), however they are still higher than ideal with regards to the amino acid components. Examination of the plots of measured concentration against predicted concentration for the model illustrates there is a tendency of the model towards over or under prediction of all the analytes of interest (**Figure 8-13**).

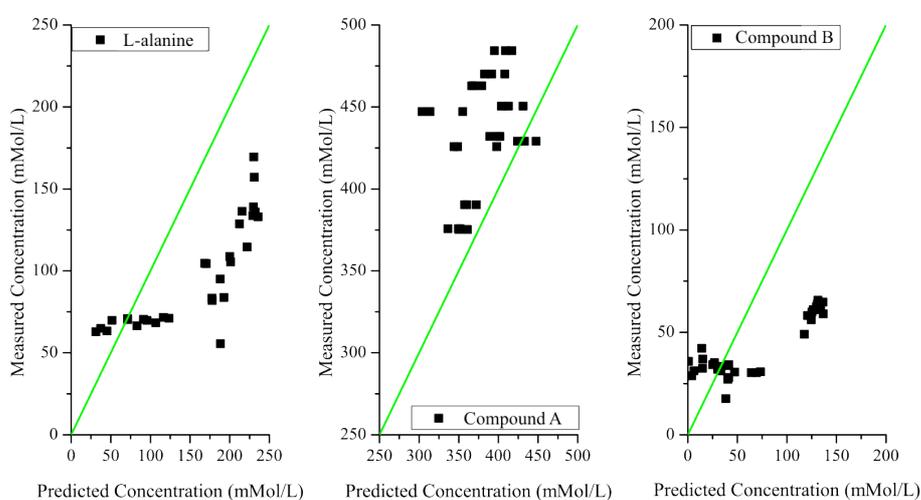


Figure 8-13 - Plots of measured concentration (mMol dm⁻³) (from HPLC methods) against predicted concentration (mMol dm⁻³) (from PLS model) for the external validation dataset.

Between the point when the spectra of the original biotransformation samples were acquired and the acquisition of spectra for the spiked biotransformation samples the ZnSe ATR crystal on the instrument was replaced. It was possible that the optics of the new ATR crystal were different from the original crystal, resulting in changes to the spectra obtained.⁸⁷

PCA was once again utilised to determine if there were any substantial differences in the spectra obtained for the original biotransformation samples and those retrospectively spiked. All mid infrared spectra had been converted to the first derivative and were mean centered. The plot of scores associated with the first principal component were plotted against the scores for the second principal component and examined.

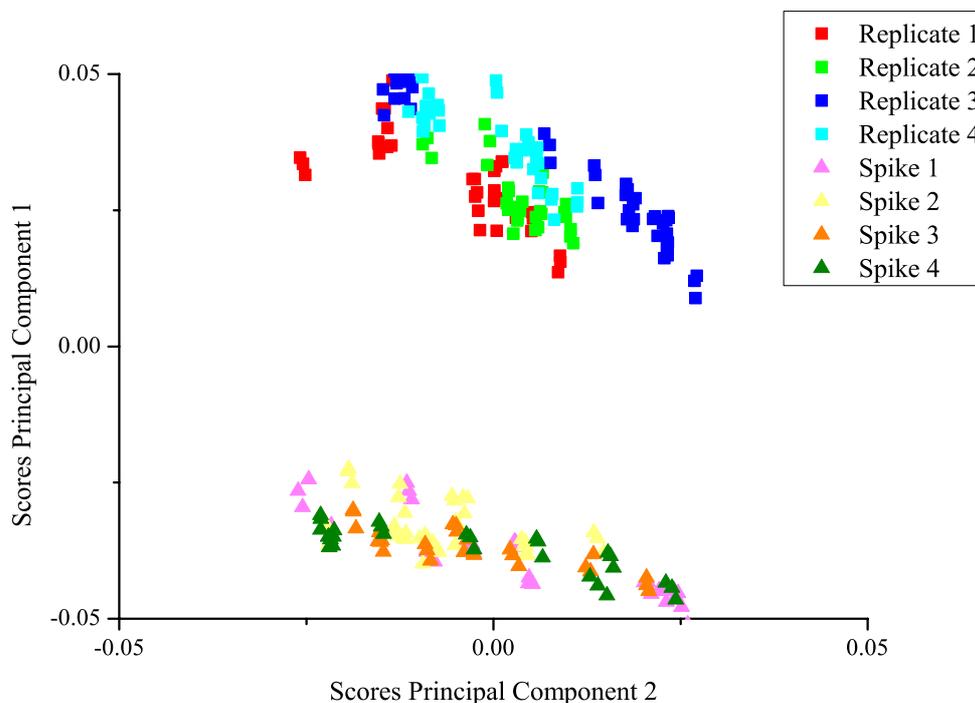


Figure 8-14 - PCA scores plot of all spectroscopic data obtained for the original samples and the retrospectively spiked samples.

The PCA scores plot clearly illustrated there were substantial spectral differences between the spectra obtained from the original samples and those samples that were

retrospectively spiked to break the co-linearity. The original samples are clustered in an area of the PCA space whilst the spiked samples occupy a completely different region.

Since the ATR crystal had been changed between acquisition of the two datasets it was concluded that the optical differences between the crystals (such as a higher light transmittance through the crystal) was responsible for this models failure to accurately predict the analyte concentrations in the validation dataset.

8.5.3 Calibration Transfer

Variations in both the near and mid infrared spectra obtained can arise as a result of either instrument fluctuations, such as instrument ageing or repair, or physical differences between instrumentation. The spectral response obtained from each infrared spectrometer is unique, even between two identical instruments from the same manufacturer.¹²² In the first case a maintenance regime can be sufficient for maintaining the predictive ability of the calibration model, however in the second situation, where multiple spectrometer systems may be employed in an industrial setting, the PLS algorithms used do not adequately model this variation.⁸⁷ There are a variety of options that could be considered to deal with this situation; individual calibration models for each instrument, the construction of a global calibration model or instrument standardisation.¹²¹

In situations where multiplexing of the spectrometer system has been employed, a similar challenge arises with variations in spectral features and response arising due to the optical differences between the various process probes.⁸⁷ These reported variations between probes are very similar to the observed differences in spectral features on changing the ATR crystal, therefore these same options for dealing with these variations should apply.

Proposing the construction of a separate calibration model for each instrument is inefficient¹²¹, and in the case of the biotransformation samples would require the

spectra of all the process samples to be re-acquired. The second suggestion of constructing a calibration model based on the spectra from both instruments, or in this case ATR crystals, was also considered however with the limited number of different crystals essentially averaging the spectral features would introduce an additional, unnecessary, error into the model.

Standardisation of the samples was therefore determined as being the most appropriate mechanism for dealing with the differences that arose between ATR crystals. The approach defines one crystal as a master with the spectra using other crystals mathematically converted to appear like the master crystal. A variety of mathematical algorithms (such as univariate correction, direct standardisation and piecewise direct standardisation) are available to achieve this transformation.¹²² Adopting this approach would require only a limited number of samples to be acquired on both crystals, but would also allow the same model to be applied should the ATR crystal require replacement again.

Ten samples from the biotransformation replicates were identified and their spectra acquired on the new instrument. Using these spectra and the original spectra of the biotransformation replicates a calibration transfer procedure was attempted. This would allow the spectra of the spiked samples acquired using one ATR crystal to be used in conjunction with the spectra of the original biotransformation samples obtained using a different ATR crystal.

8.5.3.1 Sample Selection & Identification

A total of ten of the original biotransformation samples were selected for scanning on the new ATR crystal and used to determine the parameters for the calibration transfer. Ten samples were chosen for the transfer process because this represented approximately 20% of the total (53) biotransformation samples available. Since appropriate selection of the standardisation samples was considered crucial to success, it was felt that the acquisition of approximately a fifth of the available

samples on the new crystal would provide sufficient samples that encompassed the experimental range.

Rather than selecting these samples at random using a random number table, the samples were identified using leverage and inverse calibration methods.

Leverage calibration identifies the most appropriate samples based on their distance from the multivariate mean. All the spectra from the process replicate samples were combined and using the leverage selection method the ten samples for scanning on the new crystal identified (Table 30).

Table 30 - Samples identified using leverage method to be scanned on new ATR crystal to determine the calibration transfer parameters.

Spectrum Number	Process Replicate	Sample Number	Spectrum Number	Process Replicate	Sample Number
27	2	15	34	3	7
42	3	15	10	1	10
11	1	11	16	2	4
37	3	10	8	1	8
51	4	9	38	4	11

Alternatively the selection of the samples could have been carried out using the inverse calibration method. The approach in this case was to construct a PLS model for the first analyte of the process, then calculate the inverse of this model and identify the samples based on this matrix (Table 31).

Table 31 - Samples identified using inverse method to be scanned on new ATR crystal to determine the calibration transfer parameters.

Spectrum Number	Process Replicate	Sample Number	Spectrum Number	Process Replicate	Sample Number
10	1	11*	28	3	3
25	2	15*	49	4	10
31	3	7*	4	1	4
27	3	2	41	4	2
39	3	15*	32	3	8

* Denotes sample that was also identified using leverage selection method.

A total of four of the samples were identified using both the sample selection methods. Each of these identified samples had the infrared spectrum acquired on the new ATR crystal. Samples identified from each identification method were kept separate, used to generate the transfer parameters and the results in each case compared.

The infrared spectra of the samples identified using the leverage method acquired on the original ATR crystal and the new ATR crystal were plotted to determine if there were any observable differences in the acquired spectra (Figure 8-15).

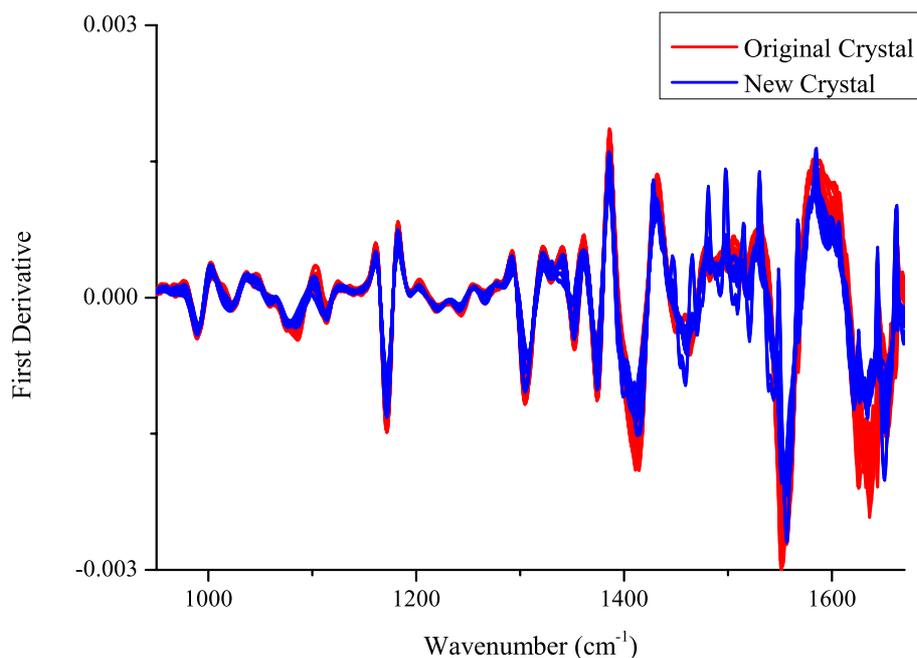


Figure 8-15 - Comparison of first derivative mid infrared spectra of the samples identified using the leverage method acquired on the original and new ATR crystals.

The overlaid plot of both spectra illustrated that there were some obvious differences between the spectra acquired using the original crystal and the new crystal. However the differences appeared mostly to be different signal intensities between the two crystals, which the proposed calibration transfer procedure should be able to correct for.

Various algorithms are available for calibration transfer procedures, in order to determine the best for this particular application three different techniques: direct standardisation (DS), piecewise direct standardisation (PDS) and spectral space transformation (SST) were evaluated. Due to the large error associated with predictions based solely on a calibration model constructed from the spiked samples (Table 28) all models were constructed from the first three replicates of the biotransformation process and the spiked samples (or the relevant transfers) and validated using the fourth replicate designated for external validation. Adopting this

approach also allowed for evaluation of the transfer in both directions (original crystal to new crystal and *vice versa*) for comparison purposes.

Each discussion of the various calibration transfer approaches has utilised the samples identified using the leverage sample selection method to determine the transfer parameter. The direction of the transfer discussed was from the original ATR crystal to the new ATR crystal. Ideally this would be the direction in which the transfer would want to be carried out for the dataset under consideration because it meant any new samples acquired could be used directly with the model and would not require any manipulation. Determination of the parameters based on the inverse sample selection method as well as transfers from the new crystal to the original were carried out, with the results detailed in Appendix III.

8.5.3.2 Direct Standardisation

Using the samples identified by the leverage method (Table 30) the transfer parameters were calculated. In the first instance the new crystal was set as the standard with the spectra acquired on the original ATR crystal converted to appear like those acquired with the new crystal.

PCA was once again utilised to identify if the procedure had been successful. If the spectra were suitably converted it would be expected that the scores for the transferred spectra would occupy the same spatial region as those of the spiked samples, which were acquired on the new crystal. The PCA plot (Figure 8-16) suggested that the calibration transfer process had been successful. Spectra acquired using the original crystal had been transformed and were now occupying the same PCA space as the spiked samples that had been acquired using the new ATR crystal.

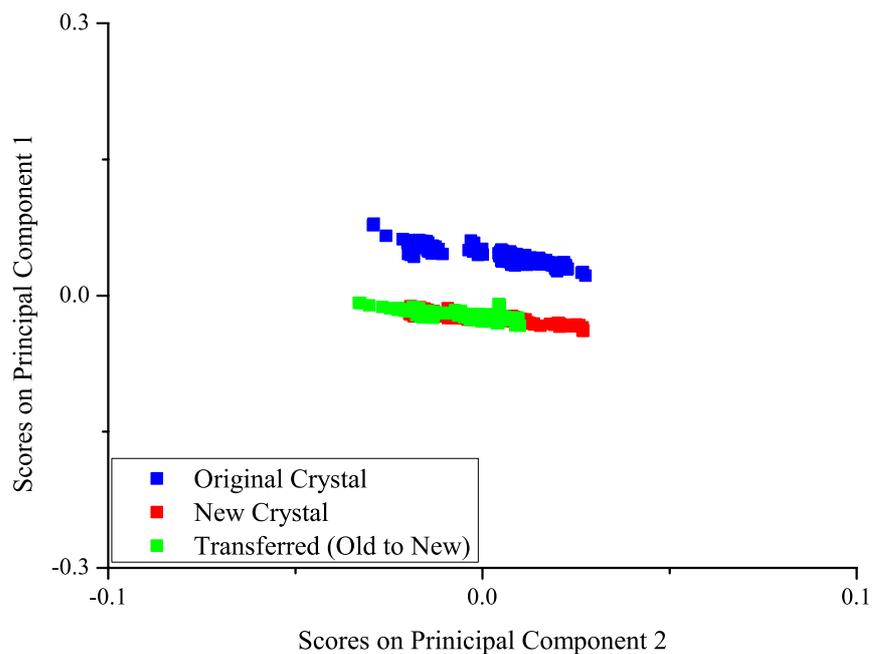


Figure 8-16 - Scores plot from PCA carried out on the original data (original ATR crystal), the spiked datasets (new ATR crystal) and the transferred spectra. Transfer parameters were determined using the samples identified using the leverage method acquired on both the original and new ATR crystal.

Based on these results the transfer parameters determined using the samples identified by the leverage method were used to convert the spectra of the process samples acquired using the original crystal to appear like those obtained from the new crystal. A PLS model for the system was constructed from the transferred spectra and spiked samples as described and tested using the spectra of the external validation dataset after having undergone the transfer process.

Table 32 - Errors associated with the prediction of concentration of the validation dataset using the direct standardisation transferred spectra.

	RMSEC (mMol dm ⁻³)	RMSECV (mMol dm ⁻³)	RMSEP (mMol dm ⁻³) [r ²]
L-alanine	39.9	41.2	85.9 [0.64]
Compound A	57.8	60.4	69.8 [0.17]
Compound B	32.4	33.7	24.4 [0.54]

By converting the spectra of the validation dataset samples using the direct standardisation approach some errors of prediction improved, specifically compound A, however higher errors of prediction were returned for some analytes compared with the model where no transfer was carried out.

Despite these indifferent results, the calibration transfer process does appear to be having a positive effect. The overlaid plot of a selection of spectra obtained on the original crystal and the same samples on the new crystal clearly indicated spectral differences between the two (Figure 8-15) which was confirmed by the PCA scores plot where spectra from both crystals occupied different spatial positions (Figure 8-14). Following the spectra undergoing transformation the PCA scores of the transformed spectra then occupied a similar spatial region (Figure 8-16) with some analytes showing an improvement in their errors of prediction.

The direct standardisation method of calibration transfer treats the spectrum as a whole, applying the transform parameters across the entire spectrum. Different methods for carrying out the transfer that consider smaller, localised regions of the spectra may be more appropriate, resulting in a more accurate transformation and therefore improved errors of calibration and prediction.

8.5.3.3 Piecewise Direct Standardisation

The piecewise direct standardisation (PDS) approach determines the transfer parameters by considering only a selected region (window) of the spectrum.⁸⁸ With

the PDS approach however it was necessary to optimise this window size to determine the optimum parameters for the calibration transfer.

Optimisation of the parameters was achieved using a script function to iteratively determine the transfer parameters based on the transfer datasets determined by either the leverage or inverse sample selection methods. In order to test the transfer parameters to determine what was the optimum window size, the leverage or inverse sample set not used to determine the transfer parameters, was used in conjunction with PCA to test the determined parameters.

Initially samples identified using the leverage selection method were utilised to generate the transfer parameters and tested using the samples identified by the inverse selection method. Transfer parameters were determined and tested with window sizes between 1 and 21 wavenumbers in increments of 2 wavenumbers. Transfers were investigated in both directions as before with the original spectra converted to the new crystal and vice versa, however to reduce future requirements in the current situation it would be preferential if the original spectra could be successfully converted to the new crystal preventing the further manipulation of any newly acquired samples.

When the calibration transfers had been completed on these small test datasets, the inverse samples acquired on both the original and new ATR crystal were added to the transferred spectra and PCA carried out to determine the optimum window size. Conversion of the spectra acquired on the original ATR crystal and converted to the new ATR crystal was investigated initially.

Although all the window sizes in the region of interest were iteratively investigated, window sizes between 1 and 5 appeared to have the most impact on the PCA scores position. In all cases, it was clear that the transformation process had been successful in converting the spectra so that the PCA scores occupied a similar region in the scores plot. Since, in this case, the spectrum of the test samples had been acquired on both crystals, the optimum window size was determined as being when the PCA

scores of the true spectra began to converge with the scores of the transferred spectra (Figure 8-17). In this particular case the window size did not appear to be having a substantial effect on the quality of the transfer procedure. The scores for a window size of 3 and 5 were identical to each other, and since these scores appeared closest to the scores of the actual samples an optimum window size of three was selected for the spectral transfer.

Having determined the optimum window size for the PDS calibration transfer, the original spectra of the biotransformation samples were subsequently converted based on the parameters determined by the samples identified using the leverage method.

Following conversion, PCA was carried out on a matrix consisting of the transformed spectra, the samples acquired using the original crystal and the spiked samples acquired using the new crystal. Evaluation of the scores associated with principal component one against those for principal component two suggested that the transfer had been successful (Figure 8-18). Although the samples were different, a successful transfer would see the PCA scores of the transferred spectra occupy the same region of the PCA plot as those of the spiked samples acquired using the new ATR crystal.

As in the case of the direct standardisation approach, the PCA scores suggested that the transfer of the spectra had been successful. Scores for the transferred spectra exhibited a substantial overlap with those of the spiked samples, which considering the bulk of the spiked samples were composed of the original samples, suggested the transfer process had been very successful.

Spectra from the fourth replicate of the biotransformation process, identified as the validation dataset, that had undergone the calibration transfer (from original crystal to new crystal) was used to predict the key analyte concentrations using the PLS model generated using the identified calibration dataset.

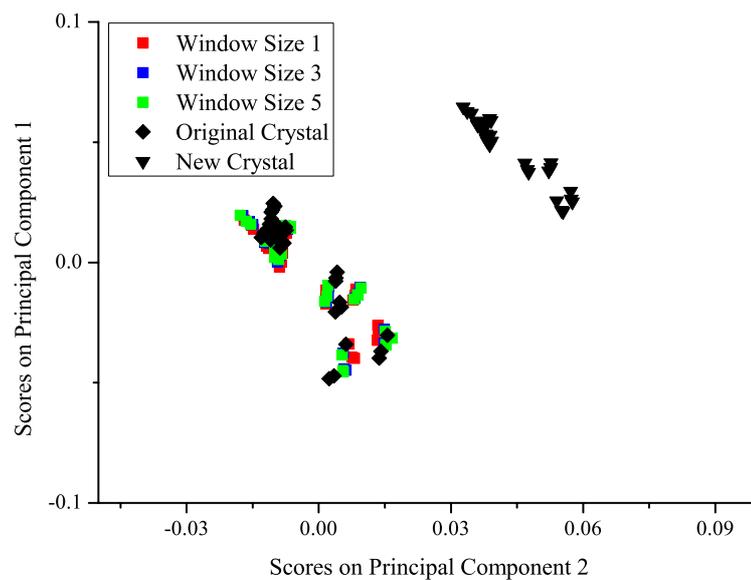


Figure 8-17 - PCA scores plot test spectra transferred using multiple window sizes to determine the optimum parameters for the calibration transfer.

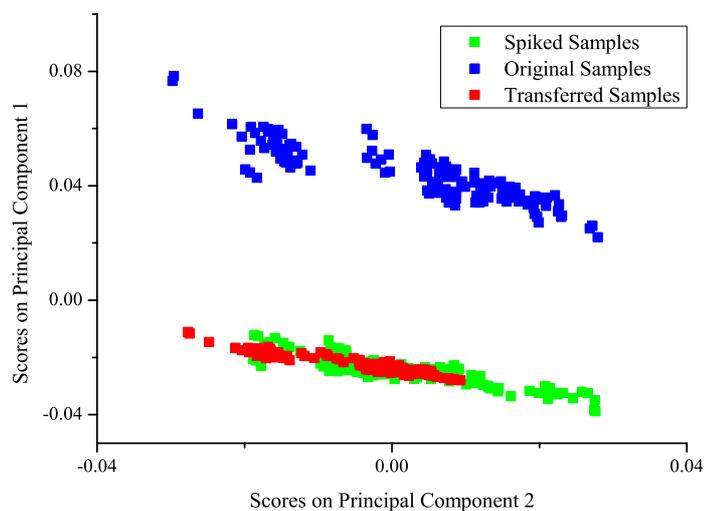


Figure 8-18 - PCA scores of original samples after calibration transfer process compared against PCA scores for the spectra acquired using the original crystal and the spiked samples acquired using the new ATR crystal.

Table 33 - Errors of calibration for the spiked PLS model and errors of prediction for the transferred spectra from the original crystal to the new crystal.

	RMSEC (mMol dm ⁻³)	RMSECV (mMol dm ⁻³)	RMSEP (mMol dm ⁻³) [r ²]
L-alanine	36.2	39.2	38.9 [0.56]
Compound A	81.4	89.0	99.7 [0.74]
Compound B	37.1	39.2	37.1 [0.63]

Compared with the errors of prediction for the validation dataset observed when the transfer was carried out using the direct standardisation method the PDS approach has demonstrated a very slight improvement for the amino acids but an increase in error for the compound A component. Overall though the errors of prediction associated with the model were still high, particularly for compound A, and further improvement could be attained.

Both the DS and PDS calibration transfer approaches are based on a very similar procedure. It was possible that the algorithm used for calculating the transfer parameters was not the most suitable. These calibration transfer algorithms, although commonly utilised, have many drawbacks, one key point where these approaches have previously failed occurred when a limited number of samples were available for acquisition on each instrument.¹²³ A scenario that was particularly relevant in the system currently under investigation.

Greater success may be achieved by utilising an alternative algorithm to determine the transfer parameters.

8.5.3.4 Spectral Space Transformation

The procedure adopted to evaluate the spectral space transformation (SST) algorithm with the in-situ biotransformation data was similar to that adopted with the PDS algorithm.

Spectra of the leverage samples acquired on both the original and new ATR crystals were used with the SST algorithm to determine the transfer parameters. These parameters were then utilised with the samples from the inverse selection method to test the calibration transfer. Spectra of the inverse samples acquired using the original ATR crystal were transferred to appear as though acquired using the new ATR crystal. PCA was utilised to compare the transferred spectra with those of the samples acquired on the new crystal and also the original data (Figure 8-19).

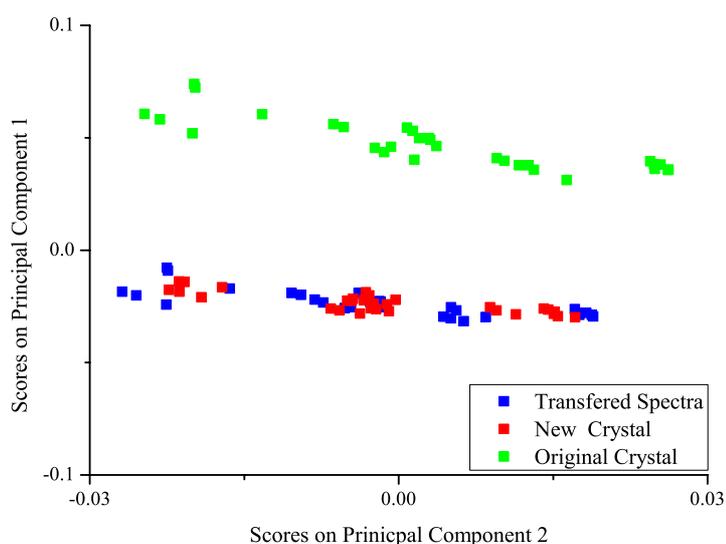


Figure 8-19 - PCA of spectra acquired using original ATR crystal, new ATR crystal and SST transfer of the original crystal to the new crystal.

The SST transfer algorithm appeared to have successfully converted the original spectra to mimic those acquired using the new crystal, evident by both the transferred spectra and these same samples acquired on the new crystal occupying the same spatial region of the PCA scores plot. Using the SST algorithm the scores of the transferred spectra appear to be better aligned with the actual samples. Previously the PCA scores plots (Figure 8-17) have indicated a transfer into the correct general region of the plot but haven't really matched up well with the scores of the actual samples. Since these were the same samples, a good calibration transfer procedure would expect to see these scores overlapping. Although some error margin was still

present, and expected, the PCA scores were clustered in the same regions as the same samples directly acquired on the new crystal.

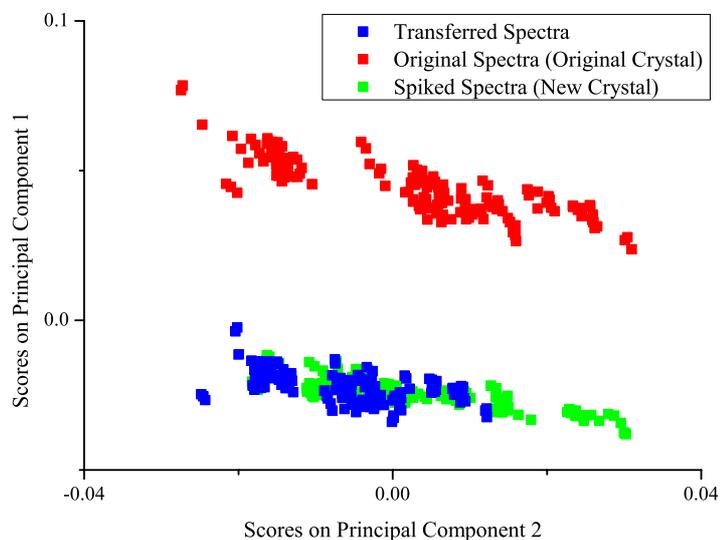


Figure 8-20 - PCA carried out on original samples acquired on original crystal, spiked samples acquired on new crystal and the transferred samples converted from the original crystal to the new crystal.

As with the previous transfer algorithms PCA suggested that the calibration transfer procedure had been successful since both the transferred spectra and spiked samples were occupying the same region in the PCA space (Figure 8-20). Using the transferred spectra for the validation dataset and the PLS model based only on the spiked samples the key analyte concentrations were predicted and compared with the reference methods.

Following on from this successful conversion of the test dataset using the SST algorithm, the original spectra of the biotransformation replicates were converted using the SST algorithm. Transfer parameters were determined using the samples identified by the leverage sample selection method with the original biotransformation samples converted to the new crystal.

The PLS model for the system was then constructed and validated using the transferred spectra of the external validation dataset.

Table 34 – Errors of calibration, cross validation and prediction for the PLS model constructed using the SST transferred spectra and spiked samples. Validation utilised the SST transferred spectra for the fourth biotransformation replicate designated for validation.

	RMSEC (mMol dm ⁻³)	RMSECV (mMol dm ⁻³)	RMSEP (mMol dm ⁻³) [r ²]
L-alanine	56.4	60.8	65.6 [0.47]
Compound A	92.6	99.5	86.5 [0.46]
Compound B	29.6	30.6	41.6 [0.58]

Use of the SST algorithm has in this case resulted in higher errors of prediction for most of the analytes of interest compared with those observed with the DS and PDS transfer methods, but also when no calibration transfer procedure was employed (Table 29). However despite the larger errors associated with model examination of the measured concentrations plotted against predicted concentration suggested that in some cases there had been an improvement to the model (Figure 8-21).

Points on these plots for both the alanine and compound B analytes appeared to be closer to the central diagonal line representing a true 1:1 relationship compared with the model where no calibration transfer procedures were employed (**Figure 8-13**).

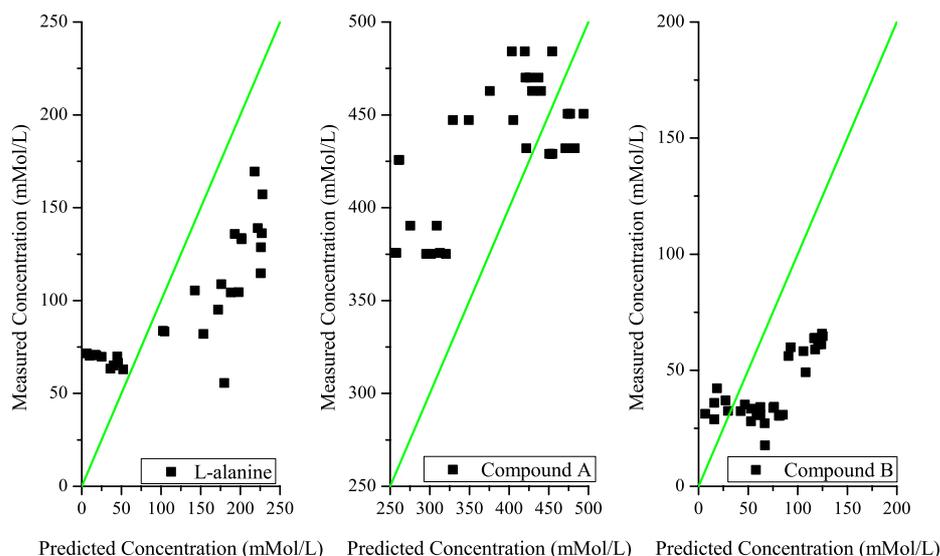


Figure 8-21 - Plot of measured concentration (mMol dm^{-3}) against predicted concentration (mMol dm^{-3}) for the validation dataset with the PLS model constructed using spectra that had been transformed from the original crystal using the SST algorithm.

Utilising the current dataset for the *in-situ* biotransformation process it was not possible to construct a robust multivariate model to quantify the key analytes in the system using infrared spectroscopy. The results obtained however, suggest there was potential to construct a suitable model. Determination of the transfer parameters was one of the key challenges associated with the transfer procedure. Suitable reference standards acquired on all ATR crystals from the outset would most likely have made the overall process much simpler.

8.5.4 Model Summary

Retrospectively, one of the key issues in the construction of a multivariate model for this system was co-linearity. Stoichiometric links within the system resulted in the spectra features of a particular analyte influencing the predicted concentration of another. Given this was a complex biological system and the nature of the process being carried out it was inevitable that these stoichiometric linkages would exist.

However in order to deal with these variations, a calibration dataset sufficiently large to ensure that all natural deviations in these linkages had occurred would be required.⁴⁹ Given that an extremely limited dataset, consisting of only four process replicates, was available for both calibration and validation there was insufficient replicates to encompass the natural variation hence the co-linearity observed in the initial model (Table 27). Upon reflection, the issue of co-linearity could have been more effectively dealt with. Perhaps making adjustments to the initial concentrations of biotransformation analytes added to the system during each replicate would have introduced sufficient variation to break these stoichiometric linkages.¹⁷

This limited number of process replicates was anticipated as being restrictive for the modelling process from the outset. Recommendations on the number of process replicates required for the construction of a robust calibration model were greater than the total number available in this case for both calibration and validation.⁸³ In combination with steps taken to deal with the co-linearity increasing the number of process replicates available for the calibration dataset would likely see a marked improvement in the quality of the resulting model.

Attempts to break the co-linearity in the dataset by a well documented technique of introducing retrospectively spiked process samples was hindered by changes in physical parameters relating to the instrumentation.³⁸ A change in ATR crystal resulted in subtle differences in the spectra between samples acquired using the original ATR crystal and a new ATR crystal. Although visually these differences appeared fairly small, PCA of spectra acquired using the original ATR crystal and the new ATR crystal occupied very different regions of a PCA plot (Figure 8-14).

Various calibration transfer algorithms, directions and parameters were investigated (Appendix III) to try and convert the spectra acquired using the original crystal to the new crystal so a calibration model could be constructed consisting of both the artificially spiked spectra and real process samples. The transfer process appeared to have been successful in that once converted the spectra acquired using the original crystal occupied the same region of a PCA plot as those acquired using a new crystal.

Whilst the transfer procedure was having a positive effect on the spectra it appeared to lack the specificity required.

One of the key requirements to achieve a successful transfer was the selection of an appropriate set of standardisation samples. These were the samples acquired on both the master instrument/crystal and slave instrument/crystals and used to calculate the transfer parameters. These standardisation samples were required to meet a number of criteria in order to result in a successful transfer procedure.

The first of these criteria was the long-term stability of the sample, in this case ideally a stability time frame in the region of months to years. Standardisation samples must be physically and chemically stable over a long time period. This was essential so that the transfer process was only correcting for instrumental variations and not true spectral variations.⁸⁸ It was assumed that the samples utilised in this situation were both chemically and physically stable. Heating the sample immediately after removal from the bioreactor to denature the enzyme should have prevented any further conversion or changes in the sample. However it was potentially possible that had the enzyme not been completely denatured some conversion may still have been occurring. In addition due to the relatively high concentrations of analytes and small volume of sample, precipitation was noted to have occurred with some of the biotransformation samples. Should this have occurred with the standardisation samples then a concentration change may have resulted introducing an additional source of error into the transfer procedure.

The second key criteria for the standardisation samples was the requirement that the selected samples were representative of the spectra to which the standardisation was to be applied. Different approaches to sample selection using the leverage and inverse sample selection methods had been proposed.⁸⁸ Both these sample selection approaches were utilised and returned similar results, which would suggest that both approaches were identifying a representative sample set.

A more successful approach may have been to use an independent standardisation sample set. Had these synthetic standards, which were known to be stable over time and covered a number of spectral features in the region of interest, been acquired on both crystals these values could have been used to generate the transfer parameters.⁸⁸

Despite various attempts to correct the spectra it was not possible to construct an independent quantitative model for this particular system. The results have demonstrated that the required steps can be taken to allow an independent model to be generated. It has also highlighted the need for a series of stable standards to be acquired on the instrumentation that will allow for the determination of a series of transfer parameters should any maintenance or changes to the system be required either during the calibration procedure or the working lifetime of the constructed model.

9. Conclusions

The aims defined at the outset of this investigation were to assess the potential of applying near and mid infrared spectroscopic techniques for the monitoring and control of industrially relevant biotransformation processes. A variety of de-racemisation and transaminase biotransformation processes were investigated and multivariate models constructed for the quantification of key analyte concentrations in each of these systems. Contributors to the constructed models were also evaluated to ensure confidence in the predictions obtained from the model.

The selective de-racemisation of racemic TBG to produce TMP and the pure enantiomer of L-TBG was investigated for potential monitoring using infrared spectroscopy. Examination of the near infrared spectrum obtained of a sample of pure TBG and TMP suggested there was insufficient variation within the acquired spectra for the construction of a reliable and robust model. Within the mid infrared region, a number of points of difference were observed in the obtained pure component spectra. A number of replicates of the process were carried out and a multivariate model constructed and externally validated based on the spectra between 900 cm^{-1} and 1620 cm^{-1} . The resulting model returned errors of prediction that were comparable with the errors of calibration and cross validation obtained for the model (4.5). Examination of the contributors to the model suggested that even although there was a stoichiometric linkage between the two analytes of interest, the resulting model was able to independently predict TBG and TMP concentrations (4.6).

The de-racemisation of racemic ABA, by conversion of D-ABA to KBA, using the D-amino acid oxidase enzyme contained within a host *Pichia pastoris* cell was also investigated as a potential process of interest for spectroscopic monitoring. With this particular process there appeared to be some subtle variations observed between the spectra obtained of ABA and KBA. Steps were taken to ensure the stoichiometric links within the system were broken from the outset by varying the initial concentrations of ABA and KBA present at the start of each biotransformation

replicate. Near infrared models were constructed based on the first overtones region, combination bands region and both of these regions. The models were then externally validated using an additional two process replicates (5.6.1). Models incorporating the combination bands region returned the best errors of prediction for the unseen process replicates. Inclusion of this region therefore restricted the technique to an at-line application as this region was unavailable when using *in-situ* near infrared. Models based on the mid infrared region were also constructed for this biotransformation process. The best model, returning the lowest errors of prediction for the validation datasets, utilised the spectroscopic regions identified by a genetic algorithm (5.7.3.2). Compared with the near infrared models the errors associated with the ABA component were lower than the comparable near infrared model. Values for KBA were marginally higher however there was a greater degree of confidence in this model due to the number of more unique spectral features observed. Examination of the regression coefficients associated with the model also suggested any issues of co-linearity within the system had been addressed.

At-line models for quantification of the two key analytes in the TBG/TMP and ABA/KBA biotransformation processes based on the mid infrared spectra had been constructed and validated. Whilst these models demonstrated the feasibility of utilising this region, ideally the *in-situ* application of the technique would be preferred from a process monitoring and control perspective. Using a diamond *in-situ* mid infrared ATR probe the de-racemisation of alanine using the D-amino acid oxidase enzymes was investigated. Models for the quantification of alanine during the biotransformation process were constructed and externally validated. This particular system returned the lowest errors of calibration, cross validation and prediction observed for any of the biotransformation processes investigated (6.5.2.3). In the case of this particular process the other analytes of interest within the system were not monitored. This was not due to limitations with the spectroscopy but due to side reactions occurring during the process involving one of the analytes of interest followed by utilisation of the product by the cells, making quantification using the developed reference methods difficult.

Investigation of the de-racemisation biotransformation processes has demonstrated that infrared spectroscopy, particularly mid infrared spectroscopy, can be utilised to monitor biotransformation processes. This represented application of the technique both at-line and *in-situ* to a challenging process both in terms of chemical complexity, due to the stoichiometric relationships between the samples and the structural similarities between the two compounds, as well as spectroscopic complexity that results from the light scattering nature of the cellular material. One of the key outcomes from these processes was the requirement to take sufficient steps to break co-linearity within the system ensuring the constructed model was able to independently quantify concentrations of the analytes of interest. This requirement was further emphasised when the transaminase class of biotransformation processes were investigated.

The transaminase biotransformation process highlighted a number of important points when attempting to use spectroscopy to monitor biotransformation processes. In common with many of the other biotransformation processes the close structural similarities between the analytes of interest meant the narrower, more specific signals found in the mid infrared region were required. Models were constructed from several replicates of the biotransformation process. However within these replicates the stoichiometric linkages were not broken by design (8.5.1). This resulted in the constructed models failing to accurately predict the analyte concentrations of the unseen validation datasets, due to co-linearity within the system. Retrospective spiking of the samples to break these stoichiometric links highlighted another important issue for the application of spectroscopic techniques to monitor bioprocesses. The nature of the bioprocess samples resulted in the degradation of the quality of ATR crystal over a period of time ultimately requiring replacement. Different crystals will have slightly different optical properties and so result in variations in the acquired spectra. These small variations in spectral features introduced a further source of error into the model thereby resulting in a failure of the model to accurately predict concentrations of the analytes of interest. Various instrument standardisation approaches were considered in an attempt to standardise the spectra from both crystals and construct a quantitative model for the system

(8.5.3). None of the algorithms utilised resulted in a calibration transfer procedure that returned the expected results. The reason behind this was most likely the samples utilised to generate the transfer parameters. Actual samples were utilised however these were of unknown stability over any time period. This highlighted the importance of acquiring the spectrum of samples that are known to be stable over a long time period. These samples can be acquired each time there was a change in crystal or maintenance operation that resulted in a change to the instruments performance and used in determining the transfer parameters.

The results presented have demonstrated the feasibility in applying infrared spectroscopic techniques as a potential method for the monitoring and control of biotransformation processes. Many of the processes investigated were still under development and not mature processes. This may have hindered the modelling process due to changes and process improvements that have occurred during the data collection exercise. Regardless of this successful models able to predict the key analyte concentrations in external validation datasets have been constructed. This has demonstrated the suitability of the approach that can then be applied to the construction of reliable and robust models for an optimised biotransformation process.

10. Future Work

Generally, many of the models constructed for the various biotransformation processes investigated were constructed based on a limited dataset. All the models presented would see an improvement in the associated errors by increasing the population of the calibration dataset.

Improvements to the method by which the amino acid content in the samples was determined would also result in an improvement to the errors of calibration and prediction of the constructed models. As the currently employed method requires a derivatisation stage, this introduces an additional error into the model that would be reduced if a direct determination method were available. Other approaches to quantifying the amino acid content of the samples such as capillary electrophoresis or liquid chromatography mass spectrometry (LC/MS) should be considered to improve the constructed models.

The model constructed for the de-racemisation of TBG biotransformation process relied heavily on the fact that variations in the initial concentrations of TBG and TMP were present in the substrate as a result of the batch preparation of substrate. Whilst this did appear to be sufficient to break co-linearity within the system a more structured approach, such as utilising a DoE to ensure independent initial concentrations of these analyte concentrations, would improve the robustness of a constructed model. The failure of the model to accurately predict the concentrations in two off-site replicates of the biotransformation process should be further investigated, to determine why this particular model failed and, if appropriate and industrially relevant, reconstruct the model so this could be achieved.

Models constructed based on the mid infrared region for the de-racemisation of ABA returned good errors of prediction for two unseen validation datasets. Attempts to predict an off-site replicate with this model appeared to start well initially but the model failed to accurately predict the latter concentrations. Further replicates of the

process should be carried out off-site to determine if this model could be used to predict the ABA and KBA concentrations of a process carried out in a different system and location. The at-line model served as a proof of concept; ideally an *in-situ* model for this process should be pursued.

Using the *in-situ* mid infrared system, a model was constructed for the quantification of alanine over the course of the biotransformation process. Side reactions within the system and the apparent utilisation of products by the cells prevented quantification of further analytes using this model. Improved reference analysis methods with lower limits of detection should be investigated in an attempt to add the ability to monitor these components to the model. Further replicates of the process should be carried out to improve the model and application of the technique for control applications also investigated.

Attempts to use pattern recognition approaches to determine the stage of a submerged culture fermentation process appeared feasible. Further replicates of the fermentation process should be carried out using the spectroscopic measurements and the artificial neural network as a means of identifying the stage of the process to try and eliminate the carbon limitation stages experienced. Overall the use of the spectroscopic approach would be expected to result in the desired biomass concentrations being reached in a shorter time frame but with the production of similar levels of enzyme following the induction phase of the process.

With the spectral data currently available there is limited scope for further progression with models for the transaminase biotransformation process. Ideally a series of stable standards would be acquired on both the original and new ATR crystal and the various transfer algorithms investigated to try and calculate a more successful set of transfer parameters. Given the damaged state of the original crystal the acquisition of such standards would not be possible. Alteration of the samples by varying the initial concentrations of the analytes of interest in the system independently of each other has proven to be an effective method of addressing co-linearity within the system. Given these difficulties it may be more appropriate to

consider acquiring the spectrum of the stable sample materials on the current ATR crystal and then retaining these samples for acquisition on a new ATR crystal when replacement of the crystal was required. Due to the optical differences between spectra in order to construct a quantitative model for the system the biotransformation process would need to be re-run this time ensuring that co-linearity had been adequately addressed.

The results presented confirmed that mid infrared spectroscopy was suitable as a method of quantifying the key analytes in these biotransformation processes both at-line and *in-situ*. The next logical progression would be to apply the technique *in-situ* to a mature process on an industrial scale.

11. References

1. D. J. Cordato, L. E. Mather and G. K. Herkes, *Journal of Clinical Neuroscience*, 2003, **10**, 649-654.
2. I. Fotheringham, I. Archer, R. Carr, R. Speight and N. J. Turner, *Biochem Soc T*, 2006, **34**, 287-290.
3. L. Pollegioni, K. Diederichs, G. Molla, S. Umhau, W. Welte, S. Ghisla and M. S. Pilone, *J. Mol. Biol.*, 2002, **324**, 535-546.
4. T. S. Hwang, H. M. Fu, L. L. Lin and W. H. Hsu, *Biotechnol. Lett.*, 2000, **22**, 655-658.
5. V. I. Tishkov and S. V. Khoronenkova, *Biochem.-Moscow*, 2005, **70**, 40-54.
6. F. D. A. U.S. Department of Health and Human Services, Rockville, MD, 2004.
7. D. Landgrebe, C. Haake, T. Hopfner, S. Beutel, B. Hitzmann, T. Scheper, M. Rhiel and K. F. Reardon, *Appl. Microbiol. Biotechnol.*, 2010, **88**, 11-22.
8. W. Chew and P. Sharratt, *Analytical Methods*, 2010, **2**, 1412-1438.
9. S. Gnoth, M. Jenzsch, R. Simutis and A. Lubbert, *J. Biotechnol.*, 2007, **132**, 180-186.
10. F. Mavituna and C. G. Sinclair, in *Practical fermentation technology*, eds. B. McNeil and M. Harvey Linda, Chichester, England ; Hoboken, NJ : Wiley, 2008, pp. 167-230.
11. A. Liese and M. Villela, *Curr. Opin. Biotechnol.*, 1999, **10**, 595-603.
12. J. Tramper, *Biotechnol. Bioeng.*, 1996, **52**, 290-295.
13. S. E. Milner and A. R. Maguire, *ARKIVOC (Gainesville, FL, U. S.)*, 2012, 321-382.
14. G. Matthews, in *Practical fermentation technology*, eds. B. McNeil and L. M. Harvey, Chichester, England ; Hoboken, NJ : Wiley, 2008.
15. L. M. Prescott, J. P. Harley and D. A. Klein, *Microbiology*, McGraw-Hill, Boston, 2002.

16. G. Macaloney, I. Draper, J. Preston, K. B. Anderson, M. J. Rollins, B. G. Thompson, J. W. Hall and B. McNeil, *Food Bioprod. Process.*, 1996, **74**, 212-220.
17. A. E. Cervera, N. Petersen, A. E. Lantz, A. Larsen and K. V. Gernaey, *Biotechnol. Prog.*, 2009, **25**, 1561-1581.
18. A. G. Cavinato, D. M. Mayes, Z. H. Ge and J. B. Callis, *Anal. Chem.*, 1990, **62**, 1977-1982.
19. S. Vaidyanathan, G. Macaloney, J. Vaughn, B. McNeil and L. M. Harvey, *Crit. Rev. Biotechnol.*, 1999, **19**, 277-316.
20. S. Vaidyanathan and B. McNeil, *Eur. Pharm. Rev.*, 1998, 43-48.
21. M. Scarff, S. A. Arnold, L. M. Harvey and B. McNeil, *Crit. Rev. Biotechnol.*, 2006, **26**, 17-39.
22. J. W. Hall and A. Pollard, 24th Annual Oak Ridge Conf on Advanced Analytical Concepts for the Clinical Laboratory - Less Invasive Technologies : From Fingersticks to Implantables, San Diego, Ca, 1992.
23. M. Navratil, A. Norberg, L. Lembren and C. F. Mandenius, *J. Biotechnol.*, 2005, **115**, 67-79.
24. J. W. Hall, B. McNeil, M. J. Rollins, I. Draper, B. G. Thompson and G. Macaloney, *Appl. Spectrosc.*, 1996, **50**, 102-108.
25. S. A. Arnold, R. Gaensakoo, L. M. Harvey and B. McNeil, *Biotechnol. Bioeng.*, 2002, **80**, 405-413.
26. G. Macaloney, J. W. Hall, M. J. Rollins, I. Draper, K. B. Anderson, J. Preston, B. G. Thompson and B. McNeil, *Bioprocess Eng.*, 1997, **17**, 157-167.
27. M. J. McShane and G. L. Cote, *Appl. Spectrosc.*, 1998, **52**, 1073-1078.
28. S. Harthun, K. Matischak and P. Friedl, *Biotechnol. Tech.*, 1998, **12**, 393-397.
29. M. J. Meurens, in *Making light work : Advances in near infrared spectroscopy*, eds. I. Murray and I. A. Cowe, V C H Publishers, New York, 1992.
30. G. Macaloney, J. W. Hall, M. J. Rollins, I. Draper, B. G. Thompson and B. McNeil, *Biotechnol. Tech.*, 1994, **8**, 281-286.
31. I. Giavasis, I. Robertson, B. McNeil and L. M. Harvey, *Biotechnol. Lett.*, 2003, **25**, 975-979.

32. S. A. Arnold, J. Crowley, S. Vaidyanathan, L. Matheson, P. Mohan, J. W. Hall, L. M. Harvey and B. McNeil, *Enzyme Microb. Technol.*, 2000, **27**, 691-697.
33. S. Vaidyanathan, A. Arnold, L. Matheson, P. Mohan, G. Macaloney, B. McNeil and L. M. Harvey, *Biotechnol. Prog.*, 2000, **16**, 1098-1105.
34. S. A. Arnold, L. Matheson, L. M. Harvey and B. McNeil, *Biotechnol. Lett.*, 2001, **23**, 143-147.
35. H. Chung, M. A. Arnold, M. Rhiel and D. W. Murhammer, *Appl. Spectrosc.*, 1996, **50**, 270-276.
36. M. R. Riley, M. Rhiel, X. J. Zhou, M. A. Arnold and D. W. Murhammer, *Biotechnol. Bioeng.*, 1997, **55**, 11-15.
37. M. R. Riley, H. M. Crider, M. E. Nite, R. A. Garcia, J. Woo and R. M. Wegge, *Biotechnol. Prog.*, 2001, **17**, 376-378.
38. M. R. Riley, M. A. Arnold, D. W. Murhammer, E. L. Walls and N. Delacruz, *Biotechnol. Prog.*, 1998, **14**, 527-533.
39. J. B. Holm-Nielsen, C. J. Lomborg, P. Oleskiewicz-Popiel and K. H. Esbensen, *Biotechnol. Bioeng.*, 2008, **99**, 302-313.
40. G. Vaccari, E. Dosi, A. L. Campi, A. Gonzalezvara, D. Matteuzzi and G. Mantovani, *Biotechnol. Bioeng.*, 1994, **43**, 913-917.
41. Z. H. Ge, A. G. Cavinato and J. B. Callis, *Anal. Chem.*, 1994, **66**, 1354-1362.
42. D. C. Hassell and E. M. Bowman, *Appl. Spectrosc.*, 1998, **52**, 18A-29A.
43. G. Steiner, in *Handbook of spectroscopy*, eds. W. R. James and C. Chemical Rubber, Cleveland : CRC Press, 1974.
44. S. A. Arnold, L. M. Harvey, B. McNeil and J. W. Hall, *Biopharm. Int.*, 2002, **15**, 26-+.
45. J. Andrews and P. Dallin, *Spectroscopy Europe*, 2003, **15**, 27-30.
46. P. Roychoudhury, R. O'Kennedy, B. McNeil and L. M. Harvey, *Anal. Chim. Acta*, 2007, **590**, 110-117.
47. N. Petersen, P. Odman, A. E. C. Padrell, S. Stocks, A. E. Lantz and K. V. Gernaey, *Biotechnol. Prog.*, 2010, **26**, 263-271.
48. E. Tamburini, G. Vaccari, S. Tosi and A. Trilli, *Appl. Spectrosc.*, 2003, **57**, 132-138.

49. S. A. Arnold, J. Crowley, N. Woods, L. M. Harvey and B. McNeil, in *Biotechnol. Bioeng.*, 2003, vol. 84, pp. 13-19.
50. N. C. Crabb and P. W. B. King, in *Process analytical chemistry*, eds. R. K. Bruce and F. McLennan, London : Blackie Academic & Professional, 1995.
51. P. Roychoudhury, B. McNeil and L. M. Harvey, *Anal. Chim. Acta*, 2007, **585**, 246-252.
52. P. Roychoudhury, L. M. Harvey and B. McNeil, *Anal. Chim. Acta*, 2006, **561**, 218-224.
53. J. Crowley, B. McCarthy, N. S. Nunn, L. M. Harvey and B. McNeil, *Biotechnol. Lett.*, 2000, **22**, 1907-1912.
54. P. Roychoudhury, L. M. Harvey and B. McNeil, *Anal. Chim. Acta*, 2006, **571**, 159-166.
55. G. Mazarevica, J. Diewok, J. R. Baena, E. Rosenberg and B. Lendl, *Appl. Spectrosc.*, 2004, **58**, 804-810.
56. D. L. Doak and J. A. Phillips, *Biotechnol. Prog.*, 1999, **15**, 529-539.
57. V. Acha, M. Meurens, H. Naveau and S. N. Agathos, *Biotechnol. Bioeng.*, 2000, **68**, 473-487.
58. Y. Raichlin and A. Katzir, *Appl. Spectrosc.*, 2008, **62**, 55A-72A.
59. H. Benson, *University physics*, New York : John Wiley, 1996.
60. P. Fayolle, D. Picque, B. Perret, E. Latrille and G. Corrieu, *Appl. Spectrosc.*, 1996, **50**, 1325-1330.
61. S. Sivakesava, J. Irudayaraj and D. Ali, *Process Biochem.*, 2001, **37**, 371-378.
62. S. Sivakesava, J. Irudayaraj and A. Demirci, *J. Ind. Microbiol. Biotechnol.*, 2001, **26**, 185-190.
63. V. G. Franco, J. C. Perin, V. E. Mantovani and H. C. Goicoeche, *Talanta*, 2006, **68**, 1005-1012.
64. J. Schenk, I. W. Marison and U. von Stockar, *J. Biotechnol.*, 2007, **128**, 344-353.
65. F. Yu, W. J. Wadsworth and J. C. Knight, *Opt. Express*, 2012, **20**, 11153-11158.
66. L. G. Grigorjeva, D. K. Millers, E. A. Kotomin, R. I. Eglitis and A. A. Lerman, *Journal of Physics D-Applied Physics*, 1996, **29**, 578-583.

67. P. Fayolle, D. Picque and G. Corrieu, *Food Control*, 2000, **11**, 291-296.
68. J. Schenk, I. W. Marison and U. von Stockar, *Anal. Chim. Acta*, 2007, **591**, 132-140.
69. D. Y. Tseng, R. Vir, S. J. Traina and J. J. Chalmers, *Biotechnol. Bioeng.*, 1996, **52**, 661-671.
70. S. A. Arnold, L. M. Harvey, B. McNeil and J. W. Hall, *Biopharm. Int.*, 2003, **16**, 47-+.
71. S. Armenta, S. Garrigues, M. de la Guardia and P. Rondeau, *Anal. Chim. Acta*, 2005, **545**, 99-106.
72. M. G. Trevisan and R. J. Poppi, *Talanta*, 2008, **75**, 1021-1027.
73. C. L. Winder, R. Cornmell, S. Schuler, R. M. Jarvis, G. M. Stephens and R. Goodacre, *Anal. Bioanal. Chem.*, 2011, **399**, 387-401.
74. A. Hayward, in *Practical fermentation technology*, eds. L. M. Harvey and MyiLibrary, Chichester, England ; Hoboken, NJ : Wiley, 2008.
75. J. G. Rosas, M. Blanco, J. M. Gonzalez and M. Alcalá, *Talanta*, 2012, **97**, 163-170.
76. J. G. Henriques, S. Buziol, E. Stocker, A. Voogd and J. C. Menezes, in *Optical sensor systems in biotechnology*, Springer-Verlag Berlin, Berlin, 2009, vol. 116, pp. 73-97.
77. J. Dahlbacka, J. Weegar, N. von Weymarn and K. Fagervik, *Biotechnol. Lett.*, 2012, **34**, 1009-1017.
78. J. McMurry, *Organic chemistry*, Sixth edn., Brooks/Cole - Thomson Learning, Belmont, CA.
79. M. R. Riley, C. D. Okeson and B. L. Frazier, *Biotechnol. Prog.*, 1999, **15**, 1133-1141.
80. B. Finn, L. M. Harvey and B. McNeil, *Yeast*, 2006, **23**, 507-517.
81. G. McLeod, K. Clelland, H. Tapp, E. K. Kemsley, R. H. Wilson, G. Poulter, D. Coombs and C. J. Hewitt, *J. Food Eng.*, 2009, **90**, 300-307.
82. D. Awotwe-Otoo, A. S. Zidan, Z. Rahman and M. J. Habib, *Aaps Pharmscitech*, 2012, **13**, 611-622.
83. C. Ruckebusch, L. Duponchel and J. P. Huvenne, *Chemometrics Intell. Lab. Syst.*, 2002, **62**, 189-198.

84. M. R. Riley, M. A. Arnold and D. W. Murhammer, *Appl. Spectrosc.*, 1998, **52**, 1339-1347.
85. S. Triadaphillou, E. Martin, G. Montague, A. Norden, P. Jeffkins and S. Stimpson, *Biotechnol. Bioeng.*, 2007, **97**, 554-567.
86. S. Gnoth, R. Simutis and A. Lubbert, *Eng. Life Sci.*, 2011, **11**, 94-106.
87. Z. P. Chen, L. J. Zhong, A. Nordon, D. Littlejohn, M. Holden, M. Fazenda, L. Harvey, B. McNeil, J. Faulkner and J. Morris, *Anal. Chem.*, 2011, **83**, 2655-2659.
88. E. Bouveresse and D. L. Massart, *Vib. Spectrosc.*, 1996, **11**, 3-15.
89. J. Schenk, I. W. Marison and U. von Stockar, 13th European Congress on Biotechnology (ECB 13), Barcelona, SPAIN, 2007.
90. J. Clayden, N. Greeves, S. Warren and P. Wothers, *Organic chemistry*, Oxford : Oxford University Press, 2001.
91. G. Strukul, *Angewandte Chemie-International Edition*, 1998, **37**, 1199-1209.
92. S. D. Doig, H. Simpson, V. Alphand, R. Furstoss and J. M. Woodley, *Enzyme Microb. Technol.*, 2003, **32**, 347-355.
93. P. A. Bird, D. C. A. Sharp and J. M. Woodley, *Org. Process Res. Dev.*, 2002, **6**, 569-576.
94. W. E. Huang, D. Hopper, R. Goodacre, M. Beckmann, A. Singer and J. Draper, *J. Microbiol. Methods*, 2006, **67**, 273-280.
95. M. R. Dadd, D. C. A. Sharp, A. J. Pettman and C. J. Knowles, *J. Microbiol. Methods*, 2000, **41**, 69-75.
96. S. Macauley-Patrick, S. A. Arnold, B. McCarthy, M. Harvey Linda and B. McNeil, *Biotechnol Lett*, 2003, **25**, 257-260.
97. S. Beutel and S. Henkel, *Appl. Microbiol. Biotechnol.*, 2011, **91**, 1493-1505.
98. R. Ulber, J. G. Frerichs and S. Beutel, *Anal. Bioanal. Chem.*, 2003, **376**, 342-348.
99. S. Kara, F. Anton, D. Solle, M. Neumann, B. Hitzmann, T. Scheper and A. Liese, *J. Mol. Catal. B-Enzym.*, 2010, **66**, 124-129.
100. M. I. Zibaii, A. Kazemi, H. Latifi, M. K. Azar, S. M. Hosseini and M. H. Ghezelaigh, *J. Photochem. Photobiol. B-Biol.*, 2010, **101**, 313-320.

101. T. Maskow, J. Lerchner, M. Peitzsch, H. Harms and G. Wolf, *J. Biotechnol.*, 2006, **122**, 431-442.
102. S. Wenda, S. Illner, A. Mell and U. Kragl, *Green Chemistry*, 2011, **13**, 3007-3047.
103. M. Yagasaki and A. Ozaki, *J. Mol. Catal. B-Enzym.*, 1998, **4**, 1-11.
104. T. D. Brock and M. T. Madigan, *Biology of microorganisms*, Pearson Prentice Hall, 2006.
105. M. Garcia-Garcia, I. Martinez-Martinez, A. Sanchez-Ferrer and F. Garcia-Carmona, *Biotechnol. Prog.*, 2008, **24**, 187-191.
106. C. J. Wei, J. C. Huang and Y. P. Tsai, *Biotechnol. Bioeng.*, 1989, **34**, 570-574.
107. W. Markle and S. Lutz, *Electrochim. Acta*, 2008, **53**, 3175-3180.
108. D. J. Ager, T. Li, D. P. Pantaleone, R. F. Senkpeil, P. P. Taylor and I. G. Fotheringham, *J. Mol. Catal. B-Enzym.*, 2001, **11**, 199-205.
109. I. G. Fotheringham, N. Grinter, D. P. Pantaleone, R. F. Senkpeil and P. P. Taylor, *Bioorg. Med. Chem.*, 1999, **7**, 2209-2213.
110. R. J. Hamilton and P. A. Sewell, *Introduction to high performance liquid chromatography*, Second edn., London Chapman and Hall, 1982.
111. B. C. Smith, *Fundamentals of fourier transform infrared spectroscopy*, Boca Raton, FL : CRC Press, 2011.
112. N. Sheppard, in *Handbook of vibrational spectroscopy*, eds. J. M. Chalmers and P. R. Griffiths, New York : J. Wiley, 2002.
113. W. Kemp, *Organic spectroscopy*, Basingstoke : Macmillan, 1987.
114. J. H. Van Der Maas, *Basic infrared spectroscopy*, Heyden & Son Ltd., London, 1969.
115. D. H. Williams and I. Fleming, *Spectroscopic methods in organic chemistry*, Fifth edn., McGraw-Hill, 1966.
116. J. M. Hollas, *Basic atomic and molecular spectroscopy*, Cambridge : Royal Society of Chemistry, 2002.
117. W. Brugel, *An introduction to infrared spectroscopy*, Methuen & Co Ltd, London, 1962.
118. J. N. Miller and J. C. Miller, *Statistics and chemometrics for analytical chemistry*, Prentice Hall, Harlow, England ; New York, 2000.

119. P. Gemperline, *Practical guide to chemometrics*, Boca Raton : Taylor&Francis, 2006.
120. B. M. Wise and B. R. Kowalski, *Process analytical chemistry*, London : Blackie Academic & Professional, 1995.
121. C. M. Gryniewicz-Ruzicka, S. Arzhantsev, L. N. Pelster, B. J. Westenberger, L. F. Buhse and J. F. Kauffman, *Appl. Spectrosc.*, 2011, **65**, 334-341.
122. E. Bouveresse, C. Casolino and C. de la Pezuela, *J. Pharm. Biomed. Anal.*, 1998, **18**, 35-42.
123. W. Du, Z. P. Chen, L. J. Zhong, S. X. Wang, R. Q. Yu, A. Nordon, D. Littlejohn and M. Holden, *Anal. Chim. Acta*, 2011, **690**, 64-70.
124. Y. F. Ge, C. L. S. Morgan, S. Grunwald, D. J. Brown and D. V. Sarkhot, *Geoderma*, 2011, **161**, 202-211.
125. J. S. Clark, P. B. Hodgson, M. D. Goldsmith and L. J. Street, *J. Chem. Soc.-Perkin Trans. 1*, 2001, 3312-3324.
126. V. J. K. Svedas, I. J. Galaev, I. L. Borisov and I. V. Berezin, *Anal. Biochem.*, 1980, **101**, 188-195.
127. J. Cazes, *Encyclopedia of chromatography*, CRC, 2004.
128. M. Roth, *Anal. Chem.*, 1971, **43**, 880-&.
129. O. S. Wong, L. A. Sternson and R. L. Schowen, *J. Am. Chem. Soc.*, 1985, **107**, 6421-6422.
130. R. M. C. Dawson, D. C. Elliott, W. H. Elliott and K. M. Jones, *Data for biochemical research*, Oxford : Clarendon Press, 1986.
131. J. Robinson, W. , E. Skelly Frame, M. and G. Frame II, M., *Undergraduate instrumental analysis*, Sixth edn., Marcel Dekker, New York, 2005.
132. D. L. Massart, B. G. M. Vandeginste, S. N. Deming, Y. Michotte and L. Kaufman, *Chemometrics : A textbook*, Elsevier, Amsterdam, 1988.
133. I. Voulgaris, PhD Thesis, University of Strathclyde, 2011.
134. N. R. Abu-Absi, B. M. Kenty, M. E. Cuellar, M. C. Borys, S. Sakhamuri, D. J. Strachan, M. C. Hausladen and Z. J. Li, *Biotechnol. Bioeng.*, 2011, **108**, 1215-1221.
135. J. J. Muller, M. Neumann, P. Scholl, L. Hilterhaus, M. Eckstein, O. Thum and A. Liese, *Anal. Chem.*, 2010, **82**, 6008-6014.

136. M. Ehly, P. J. Gemperline, A. Nordon, D. Littlejohn, J. K. Basford and M. De Cecco, *Anal. Chim. Acta*, 2007, **595**, 80-88.
137. T. S. Irvine, in *Fermentation : A practical approach*, eds. B. McNeil and L. M. Harvey, Oxford, England : IRL Press, 1990.
138. A. H. Harvey, S. G. Kaplan and J. H. Burnett, *Int. J. Thermophys.*, 2005, **26**, 1495-1514.
139. L. H. Garciarubio, *Macromolecules*, 1992, **25**, 2608-2613.
140. B. H. Junker and H. Y. Wang, *Biotechnol. Bioeng.*, 2006, **95**, 226-261.
141. V. Artyushenko, A. Bocharnikov, G. Colquhoun, C. Leach, V. Lobache, T. Sakharova and D. Savitsky, *Vib. Spectrosc.*, 2008, **48**, 168-171.
142. C. B. Minnich, P. Buskens, H. C. Steffens, P. S. Bauerlein, L. N. Butvina, L. Kupper, W. Leitner, M. A. Liauw and L. Greiner, *Org. Process Res. Dev.*, 2007, **11**, 94-97.
143. L. Kupper, H. M. Heise and L. N. Butvina, *J. Mol. Struct.*, 2001, **563**, 173-181.
144. Z. P. Chen, D. Lovett and J. Morris, *J. Process Control*, 2011, **21**, 1467-1482.
145. A. Eisenberg, J. E. Seip, J. E. Gavagan, M. S. Payne, D. L. Anton and R. DiCosimo, *J. Mol. Catal. B-Enzym.*, 1997, **2**, 223-232.
146. E. D. Guerrero, R. C. Mejias, R. N. Marin, M. P. Lovillo and C. G. Barroso, *J. Sci. Food Agric.*, 2010, **90**, 712-718.
147. H. Jiang, G. H. Liu, X. H. Xiao, S. Yu, C. L. Mei and Y. H. Ding, *Food Anal. Meth.*, 2012, **5**, 928-934.
148. J. R. Lucio-Gutierrez, J. Coello and S. Maspocho, *Food Res. Int.*, 2011, **44**, 557-565.
149. B. Lyons, in *IFPAC*, Baltimore, MD, 2010, p. 105.
150. S. L. Hennigan, J. D. Driskell, N. Ferguson-Noel, R. A. Dluhy, Y. P. Zhao, R. A. Tripp and D. C. Krause, *Appl. Environ. Microbiol.*, 2012, **78**, 1930-1935.
151. P. Negri, A. Kage, A. Nitsche, D. Naumann and R. A. Dluhy, *Chem Commun*, 2011, **47**, 8635-8637.
152. B. Lennox, G. A. Montague, A. M. Frith, C. Gent and V. Bevan, *J. Process Control*, 2001, **11**, 497-507.

153. D. M. Hawkins, *Journal of Chemical Information and Computer Sciences*, 2004, **44**, 1-12.

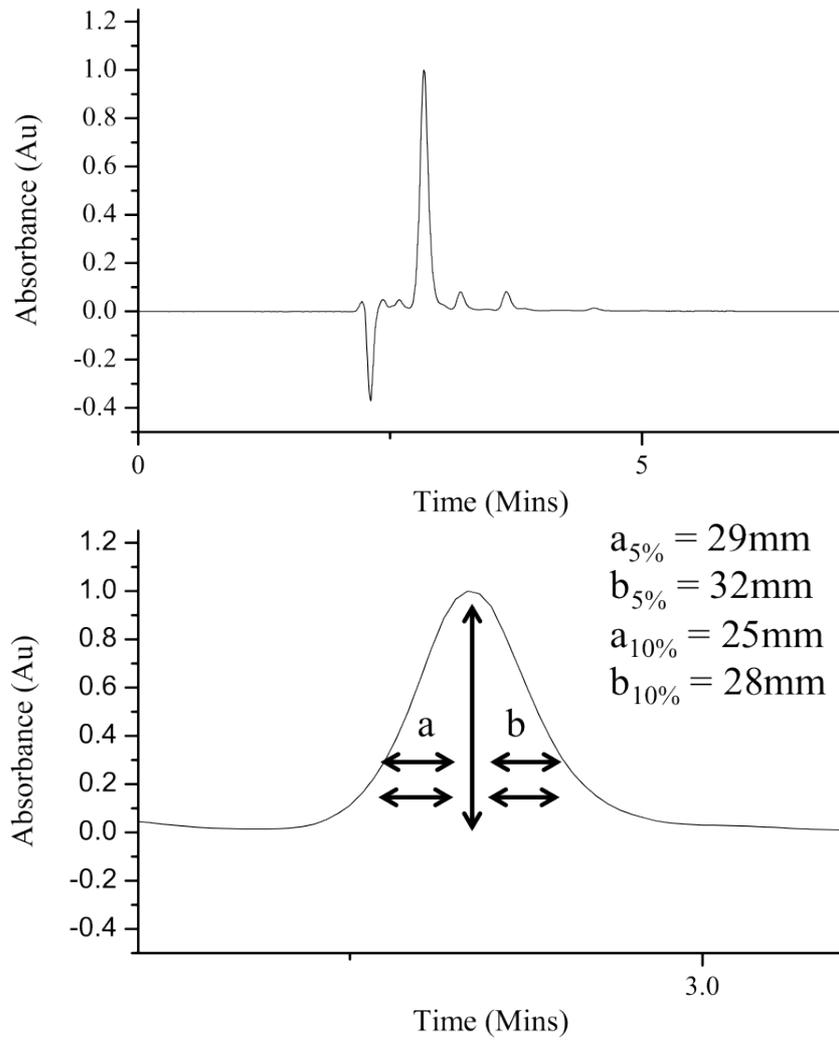
Appendix I

T distribution used in the calculation of the error associated with the reference analysis methods as obtained from Miller & Miller (2000).¹¹⁸

Degrees of Freedom (n-1)	Confidence Interval			
	90%	95%	98%	99%
1	6.31	12.71	31.82	63.66
2	2.92	4.30	6.96	9.92
3	2.35	3.18	4.54	5.84
4	2.13	2.78	3.75	4.60
5	2.02	2.57	3.36	4.03
6	1.94	2.45	3.14	3.71
7	1.89	2.36	3.00	3.50
8	1.86	2.31	2.90	3.36
9	1.83	2.26	2.82	3.25
10	1.81	2.23	2.76	3.17
12	1.78	2.18	2.68	3.05
14	1.76	2.14	2.62	2.98
16	1.75	2.12	2.58	2.92
18	1.73	2.10	2.55	2.88
20	1.72	2.09	2.53	2.85
30	1.70	2.04	2.46	2.75
50	1.68	2.01	2.40	2.68
∞	1.64	1.96	2.33	2.58

Appendix II

11.1 TMP Validation Data



Concentration	Peak Area	Predicted	(Actual-Predicted)	(Actual-Predicted) ²
1.06	15800492	1.04	-0.02	0.00
1.06	15787175	1.03	-0.03	0.00
1.06	15811747	1.04	-0.02	0.00
2.12	30663740	2.05	-0.07	0.00
2.12	30598392	2.05	-0.07	0.01
2.12	30618016	2.05	-0.07	0.00
4.24	62102028	4.20	-0.04	0.00
4.24	61823092	4.18	-0.06	0.00
4.24	61988812	4.20	-0.04	0.00
6.36	93160696	6.33	-0.03	0.00
6.36	94200664	6.40	0.04	0.00
6.36	94370192	6.41	0.05	0.00
8.48	128007288	8.71	0.23	0.05
8.48	128198096	8.73	0.25	0.06
8.48	127996000	8.71	0.23	0.05
10.6	157427744	10.73	0.13	0.02
10.6	157246736	10.71	0.11	0.01
10.6	157713872	10.74	0.14	0.02
11.66	171025840	11.66	0.00	0.00
11.66	171345472	11.68	0.02	0.00
11.66	171507904	11.69	0.03	0.00
12.72	182566736	12.45	-0.27	0.08
12.72	182940640	12.47	-0.25	0.06
12.72	182954288	12.47	-0.25	0.06
Sum (Actual-Predicted)²				0.45

Equation: $Y = 14616520x + 662782$.

$$RMSE = \sqrt{\frac{\sum(Actual - Predicted)^2}{n}}$$

$$RMSE = \sqrt{\frac{0.45}{24}} = 0.14$$

True Concentration (mMol dm⁻³)	Peak Area	Predicted Concentration (mMol dm⁻³)
8.48	128926344	8.78
8.48	128785432	8.77
8.48	128518472	8.75
8.48	128807808	8.77
8.48	129614800	8.82
8.48	129914672	8.84
8.48	129834528	8.84
8.48	129668104	8.83
8.48	129984224	8.85
8.48	129348472	8.80
Mean	129340285.6	8.80
Std Dev	538502.2624	0.04

$$\mu = \bar{x} \pm t \frac{\sigma}{\sqrt{n}}$$

$$\mu = 8.80 \pm \left(2.26 \frac{0.04}{\sqrt{10}} \right)$$

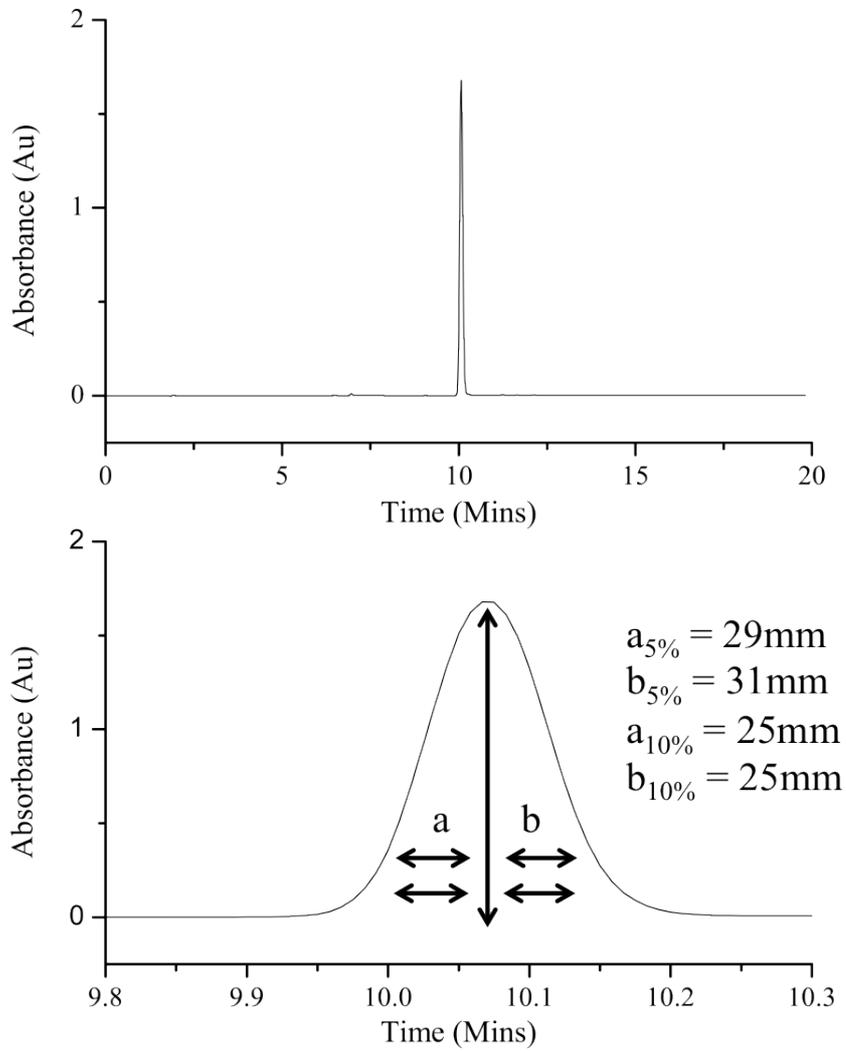
$$\mu = 8.80 \pm 0.03$$

$$\text{Total Error} = \sqrt{\text{Error 1}^2 + \text{Error 2}^2}$$

$$\text{Total Error} = \sqrt{0.14^2 + 0.03^2}$$

$$\text{Total Error} = \pm 0.14 \text{ mMol dm}^{-3}$$

11.2 ABA Validation Data



Concentration	Peak Area	Predicted	(Actual-Predicted)	(Actual-Predicted) ²
2.1	24707052	1.87	0.23	0.05
2.1	27131910	2.06	0.04	0.00
2.1	28069212	2.14	-0.04	0.00
4.2	56504872	4.37	-0.17	0.03
4.2	55012352	4.25	-0.05	0.00
4.2	53232660	4.11	0.09	0.01
6.3	80926544	6.29	0.01	0.00
6.3	82160992	6.38	-0.08	0.01
6.3	83035872	6.45	-0.15	0.02
8.4	106246752	8.28	0.12	0.02
8.4	108341344	8.44	-0.04	0.00
8.4	106359192	8.28	0.12	0.01
10.5	137662432	10.74	-0.24	0.06
10.5	137658656	10.74	-0.24	0.06
10.5	132861632	10.37	0.13	0.02
12.6	160671632	12.55	0.05	0.00
12.6	164573952	12.86	-0.26	0.07
12.6	155127824	12.11	0.49	0.24
Sum (Actual-Predicted)²				0.59

Equation: $Y=12732193x + 878652$

$$RMSE = \sqrt{\frac{0.59}{18}} = 0.18$$

True Concentration (mMol dm⁻³)	Peak Area	Predicted Concentration (mMol dm⁻³)
5.25	70905120	5.50
5.25	73146960	5.68
5.25	70994912	5.51
5.25	71366744	5.54
5.25	72031424	5.59
5.25	73032376	5.67
5.25	72790480	5.65
5.25	72105920	5.59
5.25	73433824	5.70
5.25	70190456	5.44
Mean	71999821.6	5.59
Std Dev	1103462.243	0.09

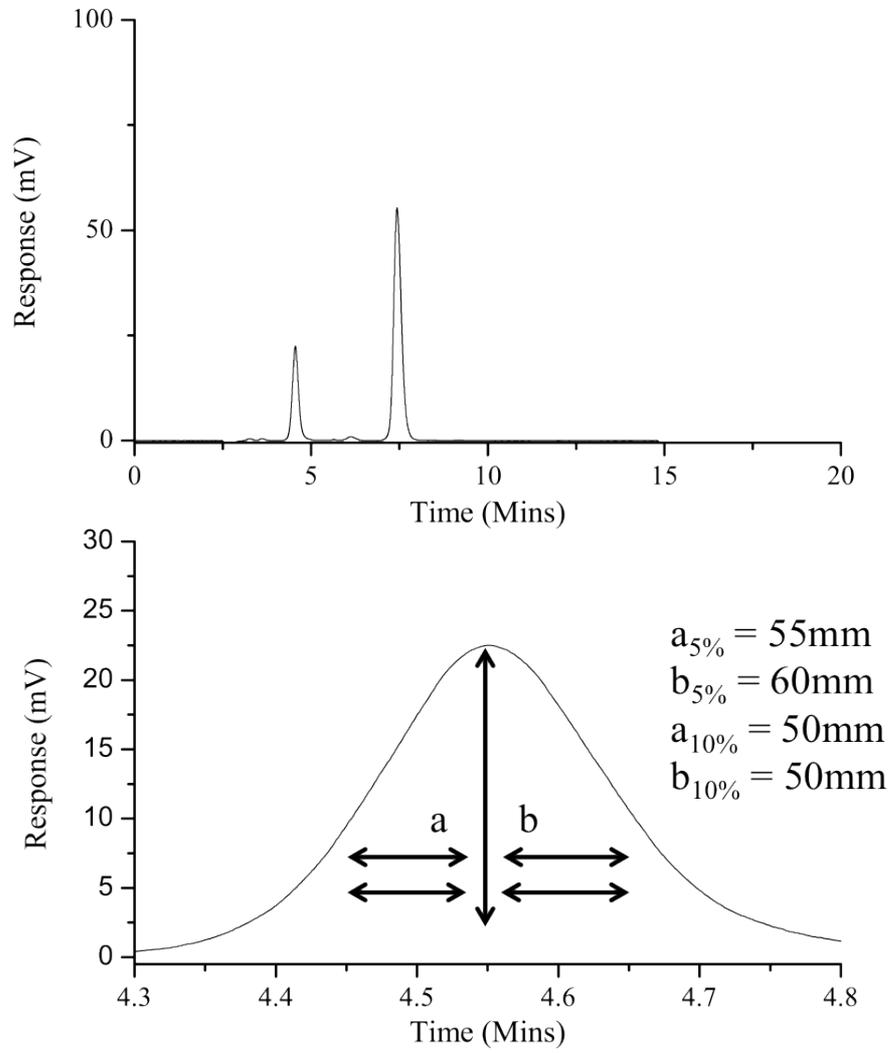
$$\mu = 5.59 \pm \left(2.26 \frac{0.09}{\sqrt{10}} \right)$$

$$\mu = 5.59 \pm 0.06$$

$$Total Error = \sqrt{0.18^2 + 0.06^2}$$

$$Total Error = \pm 0.19 \text{ mMol dm}^{-3}$$

11.3 Keto Butyric Acid



Concentration	Peak Area	Predicted	(Actual - Predicted)	(Actual - Predicted) ²
0.6	15620	0.55	0.05	0.00
0.6	16855	0.59	0.01	0.00
0.6	16069	0.56	0.04	0.00
1.2	35575	1.16	0.04	0.00
1.2	36195	1.18	0.02	0.00
1.2	36299	1.18	0.02	0.00
2.4	78544	2.47	-0.07	0.00
2.4	79692	2.50	-0.10	0.01
2.4	78841	2.47	-0.07	0.01
4.8	155900	4.82	-0.02	0.00
4.8	154922	4.79	0.01	0.00
4.8	156339	4.84	-0.04	0.00
7.2	229696	7.07	0.13	0.02
7.2	231285	7.12	0.08	0.01
7.2	231795	7.13	0.07	0.00
9.6	317037	9.73	-0.13	0.02
9.6	313383	9.62	-0.02	0.00
9.6	315103	9.67	-0.07	0.01
12.0	393633	12.06	-0.06	0.00
12.0	394117	12.08	-0.08	0.01
12.0	393485	12.06	-0.06	0.00
13.2	427732	13.10	0.10	0.01
13.2	428494	13.13	0.07	0.01
13.2	428420	13.12	0.08	0.01
Sum (Actual-Predicted) ²				0.11

Equation: $Y = 32824x - 2376.9$

$$RMSE = \sqrt{\frac{0.11}{24}} = 0.07$$

True Concentration	Peak Area	Predicted Concentration
6	200998	6.195915023
6	201571	6.213371743
6	198389	6.116430585
6	201825	6.221109975
6	199669	6.1554264
6	200134	6.169592848
6	200774	6.189090755
6	202245	6.233905476
6	200272	6.173797084
6	201542	6.212488244
	Mean	6.188112813
	Std. Dev	0.035330911

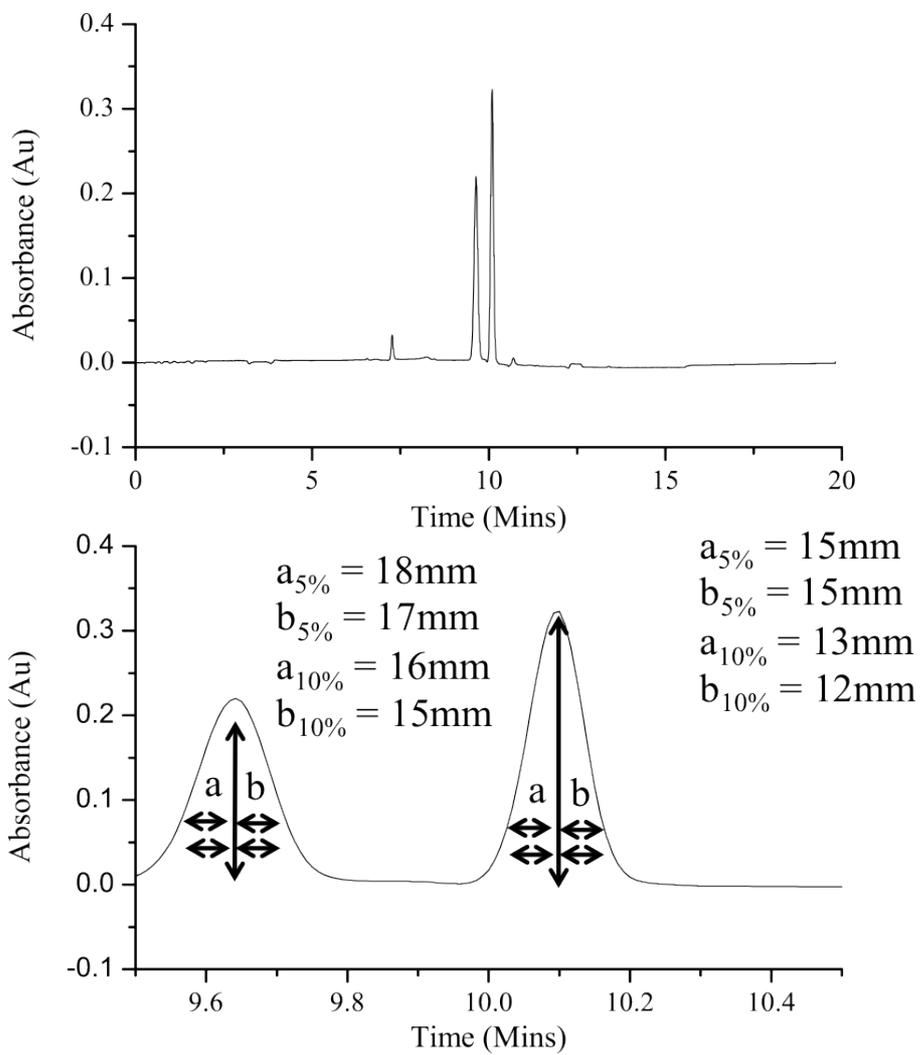
$$\mu = 6.19 \pm \left(2.26 \frac{0.03}{\sqrt{10}} \right)$$

$$\mu = 6.19 \pm 0.02$$

$$\text{Total Error} = \sqrt{0.07^2 + 0.02^2}$$

$$\text{Total Error} = \pm 0.07 \text{ mMol dm}^{-3}$$

11.4 L-alanine & Compound B Validation Data.



L-Alanine

Concentration	Peak Area	Predicted	(Actual - Predicted)	(Actual - Predicted)²
2.47	23333426	2.58	-0.11	0.01
2.47	24915686	2.73	-0.26	0.07
2.47	24036506	2.65	-0.18	0.03
3.71	35964464	3.77	-0.06	0.00
3.71	35862616	3.76	-0.05	0.00
3.71	34802784	3.66	0.05	0.00
4.94	46343064	4.74	0.20	0.04
4.94	46360780	4.74	0.20	0.04
4.94	47470852	4.84	0.10	0.01
6.18	60066884	6.02	0.16	0.03
6.18	60330764	6.04	0.14	0.02
6.18	61294136	6.13	0.05	0.00
8.65	84360272	8.29	0.36	0.13
8.65	86162672	8.46	0.19	0.04
8.65	88259024	8.66	-0.01	0.00
9.88	103652072	10.10	-0.22	0.05
9.88	105487040	10.27	-0.39	0.15
9.88	103304544	10.06	-0.18	0.03
Sum (Actual - Predicted)²				0.66

Compound B

Concentration	Peak Area	Predicted	(Actual - Predicted)	(Actual - Predicted)²
2.18	30053402	2.05	0.13	0.02
2.18	32510218	2.23	-0.05	0.00
2.18	31356096	2.15	0.03	0.00
3.27	47770312	3.36	-0.09	0.01
3.27	47591276	3.35	-0.08	0.01
3.27	45508780	3.19	0.08	0.01
4.36	59092804	4.19	0.17	0.03
4.36	59178840	4.20	0.16	0.03
4.36	62030292	4.41	-0.05	0.00
5.45	78075936	5.60	-0.15	0.02
5.45	76407976	5.47	-0.02	0.00
5.45	78122688	5.60	-0.15	0.02
7.63	107650400	7.78	-0.15	0.02
7.63	108132040	7.82	-0.19	0.04
7.63	106080152	7.67	-0.04	0.00
8.72	119837872	8.68	0.04	0.00
8.72	119347544	8.65	0.07	0.01
8.72	116424944	8.43	0.29	0.08
Sum (Actual - Predicted)²				0.29

Equation (L-alanine): $Y = 10693842 x - 4304087$

Equation (Compound B): $Y = 13534860 x + 2314483$

$$RMSE = \sqrt{\frac{0.66}{18}} = 0.19$$

$$RMSE = \sqrt{\frac{0.29}{18}} = 0.13$$

Injection	L-Alanine Concentration	L-Alanine Predicted	Compound B Concentration	Compound B Predicted
1	4.94	5.26	4.36	4.48
2	4.94	5.18	4.36	4.38
3	4.94	5.21	4.36	4.46
4	4.94	5.04	4.36	4.14
5	4.94	5.04	4.36	4.22
6	4.94	5.13	4.36	4.31
7	4.94	5.13	4.36	4.25
8	4.94	4.98	4.36	3.96
9	4.94	5.03	4.36	4.01
10	4.94	5.14	4.36	4.19
	Mean	5.11	Mean	4.24
	Std Dev	0.09	Std Dev	0.18

$$\mu = 5.11 \pm \left(2.26 \frac{0.09}{\sqrt{10}}\right)$$

$$\mu = 5.11 \pm 0.06$$

$$\mu = 4.24 \pm \left(2.26 \frac{0.18}{\sqrt{10}}\right)$$

$$\mu = 4.24 \pm 0.13$$

$$Total Error = \sqrt{Error 1^2 + Error 2^2}$$

$$Total Error = \sqrt{0.19^2 + 0.06^2}$$

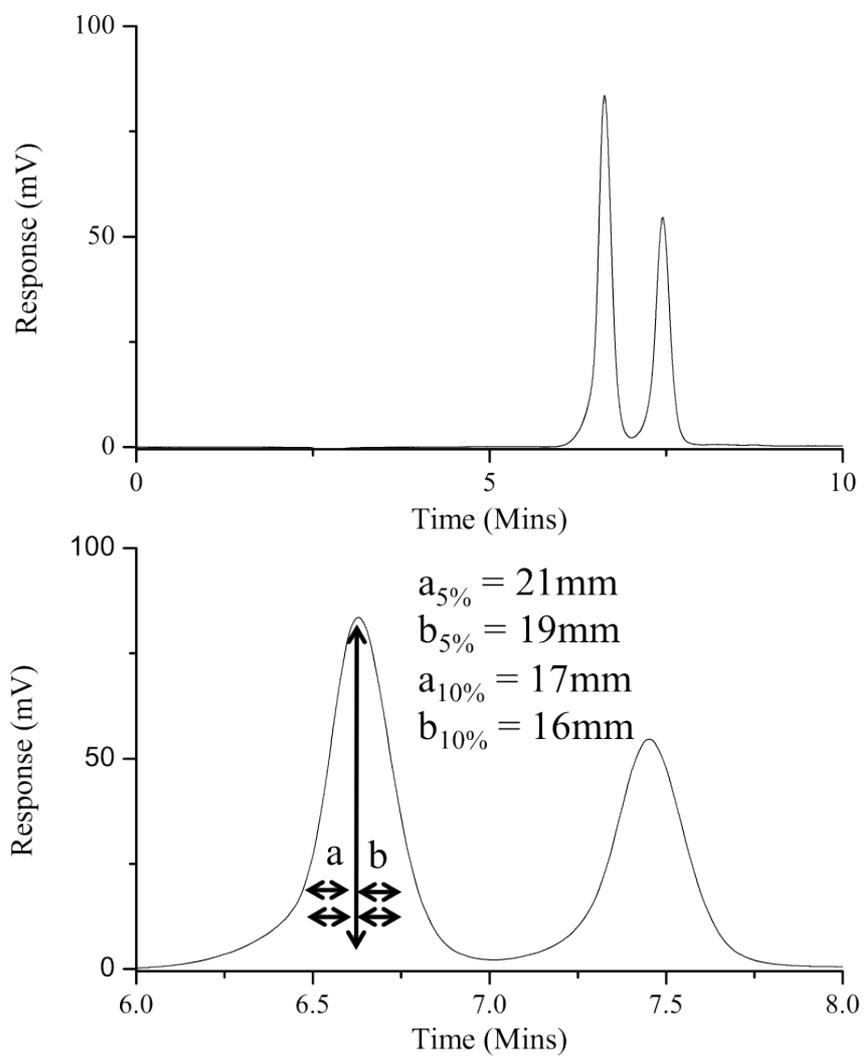
$$Total Error = \pm 0.20 \text{ mMol dm}^{-3}$$

$$Total Error = \sqrt{Error 1^2 + Error 2^2}$$

$$Total Error = \sqrt{0.13^2 + 0.13^2}$$

$$Total Error = \pm 0.18 \text{ mMol dm}^{-3}$$

11.5 Compound A



Concentration	Peak Area	Predicted	(Actual - Predicted)	(Actual - Predicted) ²
5.2	142528	5.60	-0.40	0.16
5.2	145753	5.70	-0.50	0.25
5.2	143905	5.64	-0.44	0.20
13	374602	12.93	0.07	0.00
13	374075	12.92	0.08	0.01
13	375696	12.97	0.03	0.00
19.5	571608	19.16	0.34	0.12
19.5	568345	19.06	0.44	0.20
19.5	564945	18.95	0.55	0.30
26	759317	25.09	0.91	0.83
26	759626	25.10	0.90	0.81
26	765345	25.28	0.72	0.52
39	1234751	40.12	-1.12	1.24
39	1235194	40.13	-1.13	1.28
39	1235219	40.13	-1.13	1.28
65	1998051	64.24	0.76	0.58
65	2021668	64.98	0.02	0.00
65	2025243	65.09	-0.09	0.01
Sum (Actual - Predicted)²				<u>7.78</u>

Equation: $Y = 31645x - 34721$.

$$RMSE = \sqrt{\frac{7.78}{18}} = 0.66$$

True Concentration	Peak Area	Predicted Concentration
32.5	1033972	33.77
32.5	1037947	33.90
32.5	1033828	33.77
32.5	1033390	33.75
32.5	1021910	33.39
32.5	1021624	33.38
32.5	1014400	33.15
32.5	1014791	33.16
32.5	1016296	33.21
32.5	1011872	33.07
Mean		33.46
Std Dev		0.31

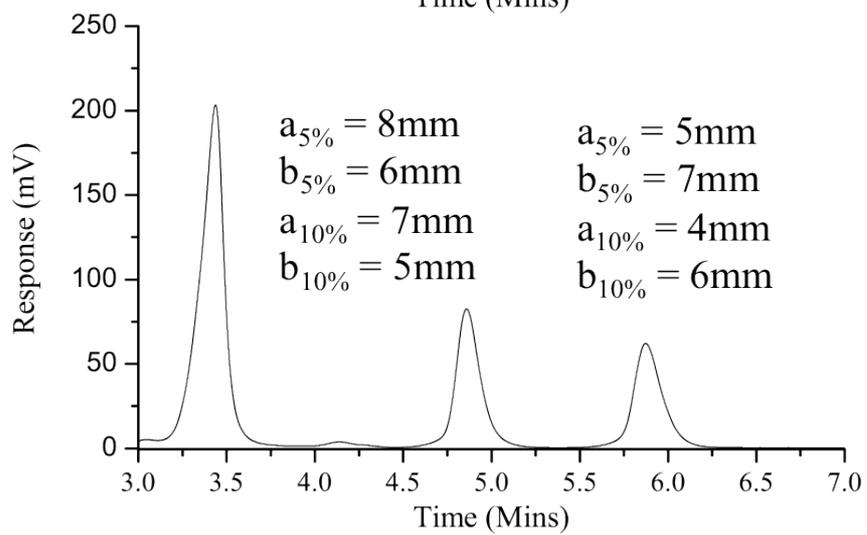
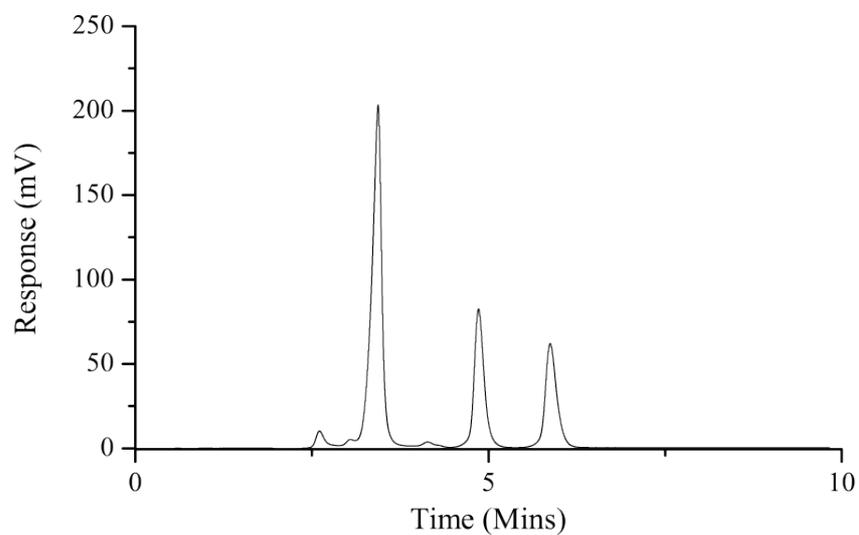
$$\mu = 33.46 \pm \left(2.26 \frac{0.31}{\sqrt{10}} \right)$$

$$\mu = 33.46 \pm 0.22$$

$$\textit{Total Error} = \sqrt{0.66^2 + 0.22^2}$$

$$\textit{Total Error} = \pm 0.70 \text{ mMol dm}^{-3}$$

11.6 Pyruvic Acid & Acetic Acid



Pyruvic Acid

Concentration	Peak Area	Predicted	(Actual - Predicted)	(Actual - Predicted) ²
16.52	335050	19.33	-2.81	7.88
16.52	335956	19.37	-2.85	8.15
16.52	344579	19.82	-3.30	10.89
41.3	515533	28.67	12.63	159.45
41.3	517648	28.78	12.52	156.70
41.3	517454	28.77	12.53	156.95
82.6	1693307	89.66	-7.06	49.81
82.6	1692240	89.60	-7.00	49.03
82.6	1684831	89.22	-6.62	43.81
123.9	2440417	128.34	-4.44	19.74
123.9	2447000	128.68	-4.78	22.88
123.9	2463243	129.52	-5.62	31.64
165.2	3266828	171.13	-5.93	35.21
165.2	3263171	170.94	-5.74	33.00
165.2	3269051	171.25	-6.05	36.59
206.5	3811041	199.31	7.19	51.65
206.5	3796538	198.56	7.94	63.01
206.5	3767867	197.08	9.42	88.78
Sum (Actual - Predicted) ²				1025.16

Acetic Acid

Concentration	Peak Area	Predicted	(Actual - Predicted)	(Actual - Predicted) ²
16.92	134262.00	19.40	-2.48	6.14
16.92	132449.00	19.15	-2.23	4.96
16.92	136432.00	19.70	-2.78	7.72
42.30	211591.00	30.12	12.18	148.36
42.30	210668.00	29.99	12.31	151.49
42.30	211670.00	30.13	12.17	148.09
84.60	648127.00	90.65	-6.05	36.55
84.60	645879.00	90.33	-5.73	32.88
84.60	647443.00	90.55	-5.95	35.41
126.90	948714.00	132.32	-5.42	29.40
126.90	947765.00	132.19	-5.29	27.99
126.90	947810.00	132.20	-5.30	28.06
169.20	1281727.00	178.49	-9.29	86.39
169.20	1280264.00	178.29	-9.09	82.66
169.20	1276733.00	177.80	-8.60	74.00
211.50	1452193.00	202.13	9.37	87.80
211.50	1447039.00	201.42	10.08	101.70
211.50	1432482.00	199.40	12.10	146.48
Sum (Actual - Predicted) ²				1236.09

Equation (Pyruvic Acid): $Y = 19312x - 38210$

Equation (Acetic Acid): $Y = 7212x - 5644$

$$MSE = \sqrt{\frac{1025.16}{18}} = 7.55$$

$$RMSE = \sqrt{\frac{1236.09}{18}} = 8.30$$

Replicate	Peak Area	Concentration	Peak Area	Concentration
1	3942646	206.13	1550965	215.82
2	3946789	206.34	1553671	216.20
3	3940474	206.02	1550228	215.72
4	3952378	206.63	1557491	216.73
5	3948105	206.41	1555188	216.41
6	4006909	209.46	1536681	213.84
7	3948366	206.42	1554851	216.36
8	3953829	206.71	1552082	215.98
9	3925736	205.25	1552491	216.04
10	3945901	206.30	1549923	215.68
	Mean	206.57	Mean	215.88
	Std. Dev.	1.09	Std. Dev.	0.79

$$\mu = 206.57 \pm \left(2.26 \frac{1.09}{\sqrt{10}}\right)$$

$$\mu = 206.57 \pm 0.78$$

$$\mu = 215.88 \pm \left(2.26 \frac{0.79}{\sqrt{10}}\right)$$

$$\mu = 215.88 \pm 0.56$$

$$Total\ Error = \sqrt{7.55^2 + 0.78^2}$$

$$Total\ Error = \pm 7.60\ \text{mMol dm}^{-3}$$

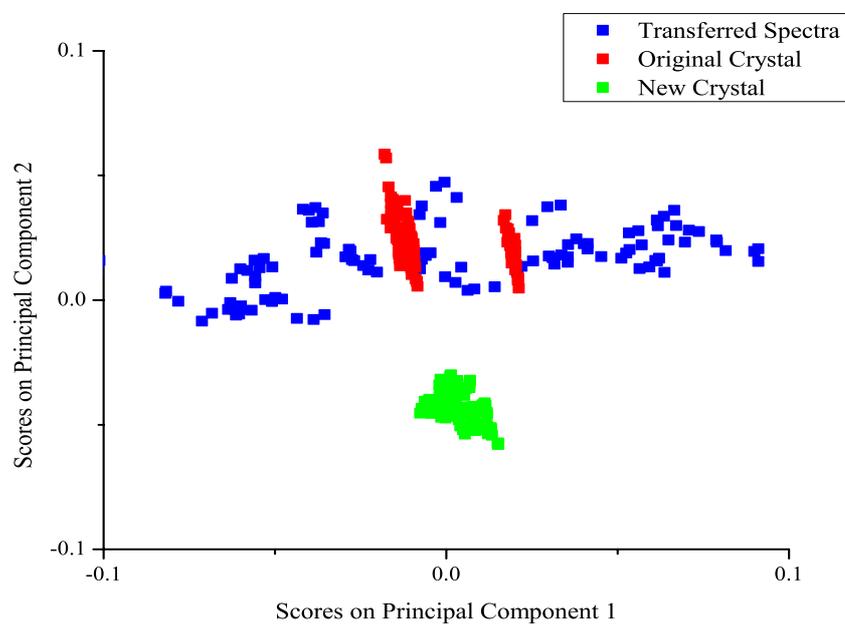
$$Total\ Error = \sqrt{8.30^2 + 0.56^2}$$

$$Total\ Error = \pm 8.31\ \text{mMol/}$$

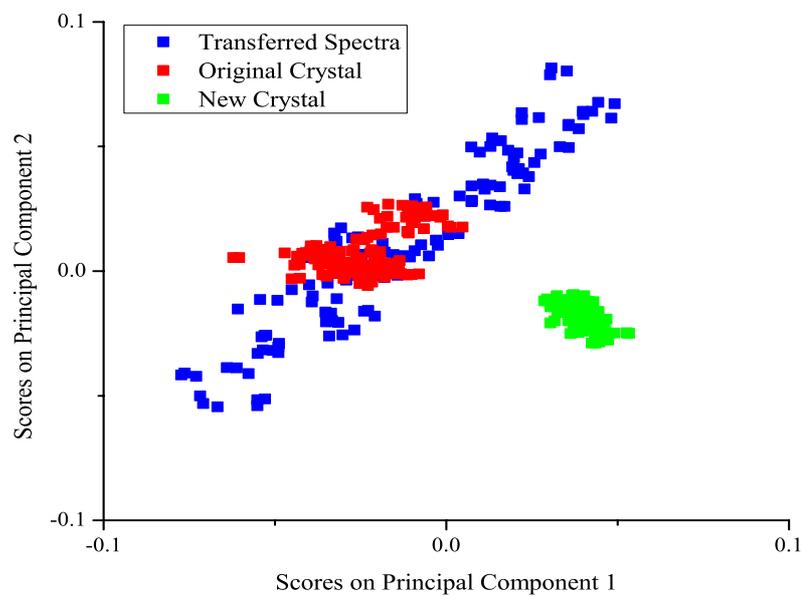
Appendix III

Sample	L-Alanine Spike (mMol/L)	Compound A Spike (mMol/L)	Compound B Spike (mMol/L)	Biotransformation Sample
1	113	0	16.8	ISB2-03
2	90	260	0	ISB2-04
3	113	0	0	ISB2-05
4	0	260	0	ISB2-06
5	45	260	77	ISB2-07
6	113	209	77	ISB2-08
7	90	104	0	ISB2-10
8	0	260	16.8	ISB2-11
9	45	0	27.6	ISB2-12
10	0	104	27.6	ISB3-01
11	113	260	27.6	ISB3-02
12	45	209	27.6	ISB3-03
13	90	260	77	ISB3-04
14	90	209	16.8	ISB3-05
15	90	0	0	ISB2-02
16	45	104	77	ISB3-06
17	0	0	27.6	ISB3-07
18	113	209	27.6	ISB3-08
19	90	104	16.8	ISB3-09
20	0	209	0	ISB3-10
21	45	0	0	ISB2-01
22	0	209	16.8	ISB3-11
23	45	104	16.8	ISB3-12
24	113	104	27.6	ISB3-14
25	0	0	77	ISB4-01
26	90	0	27.6	ISB4-02
27	0	104	77	ISB4-03
28	113	260	16.8	ISB4-04
29	90	104	77	ISB4-05
30	90	209	27.6	ISB4-06
31	113	209	0	ISB4-07
32	45	260	0	ISB4-08
33	113	104	0	ISB4-09
34	45	0	16.8	ISB5-01
35	113	0	77	ISB5-02
36	45	260	27.6	ISB5-03
37	113	260	77	ISB5-04
38	0	104	0	ISB5-05
39	0	0	16.8	ISB5-06
40	45	209	16.8	ISB5-07
41	113	104	16.8	ISB5-08
42	90	209	77	ISB5-11

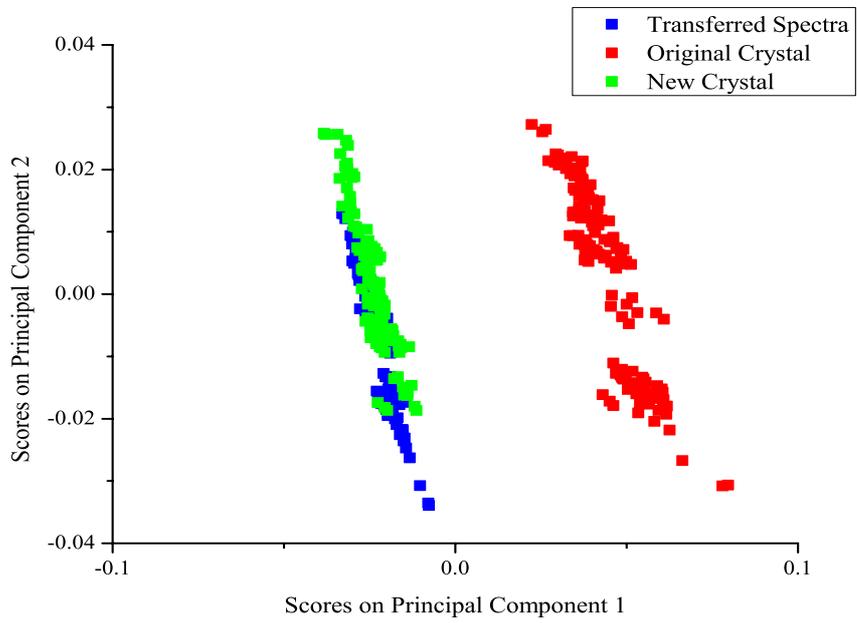
Direct standardisation, leverage sample selection new crystal to original crystal.



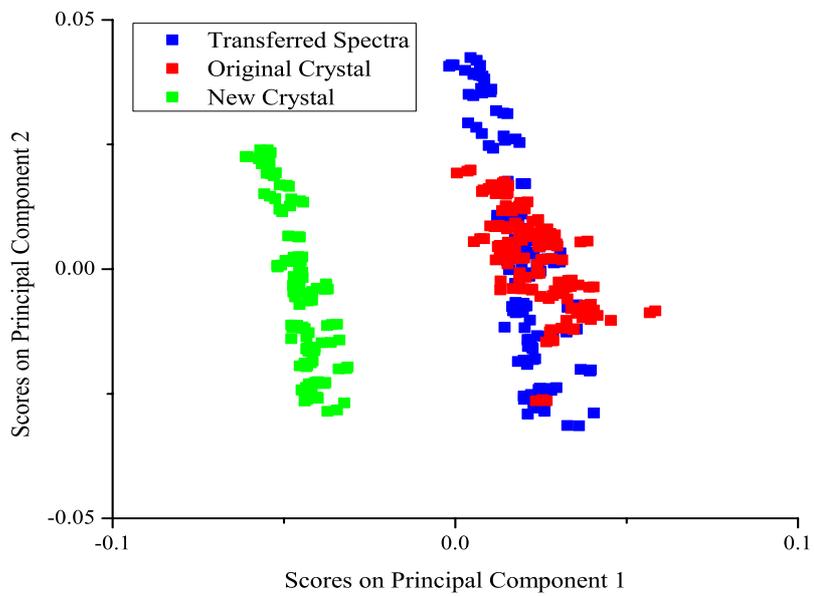
Direct standardisation, inverse sample selection new crystal to original crystal.



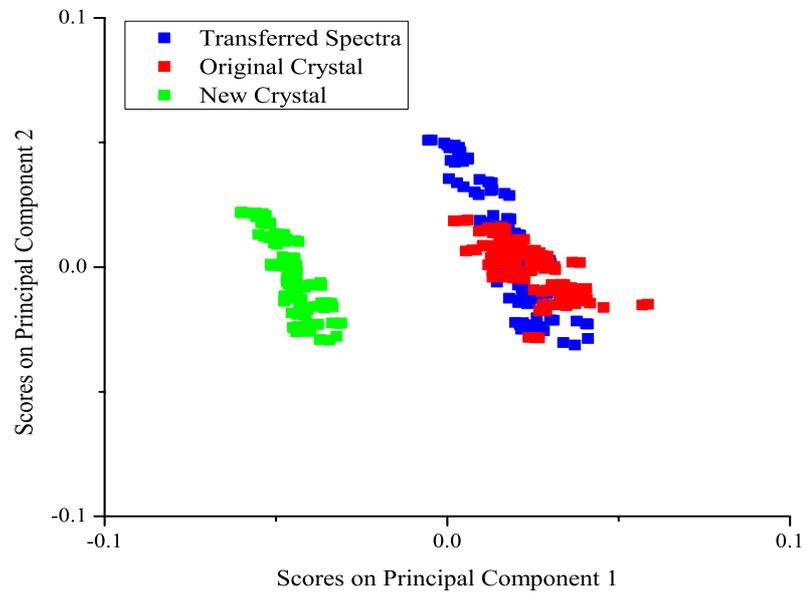
Direct standardisation, inverse sample selection original crystal to new crystal.



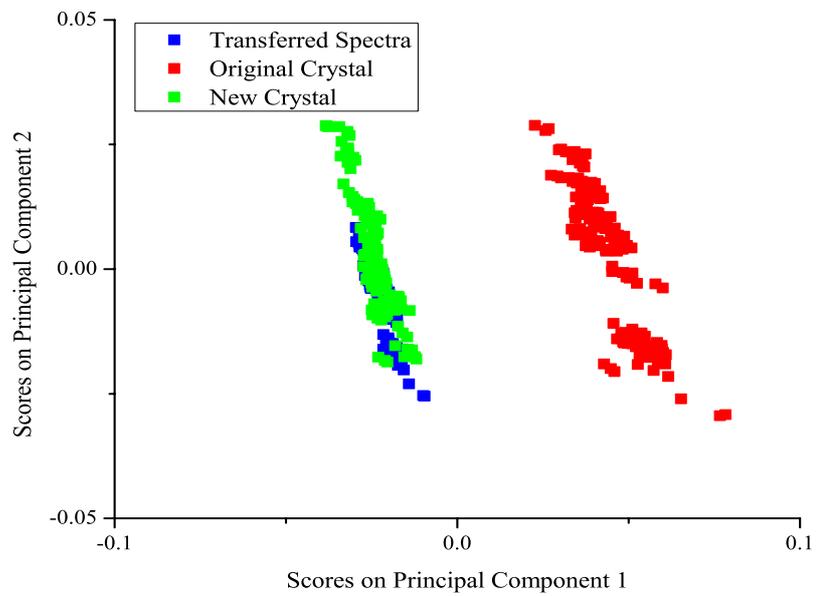
PDS, leverage sample selection new crystal to original crystal



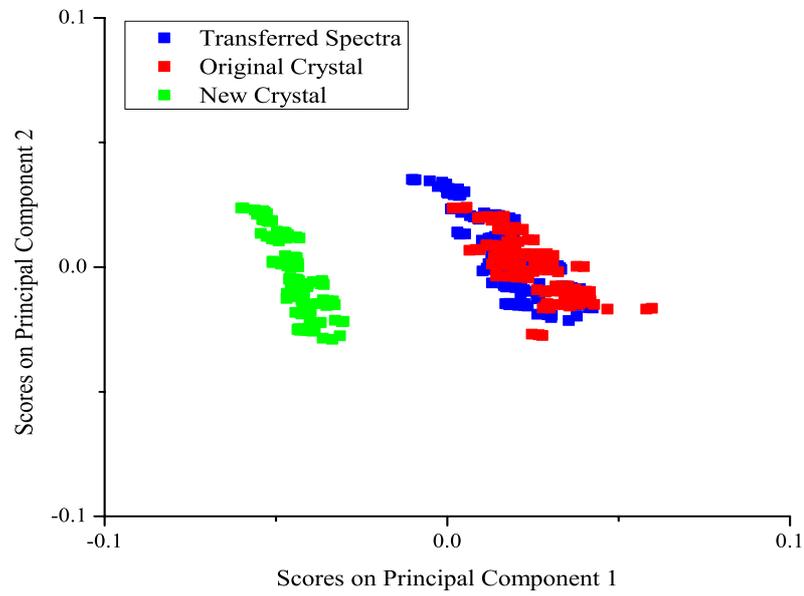
PDS, inverse sample selection new crystal to original crystal



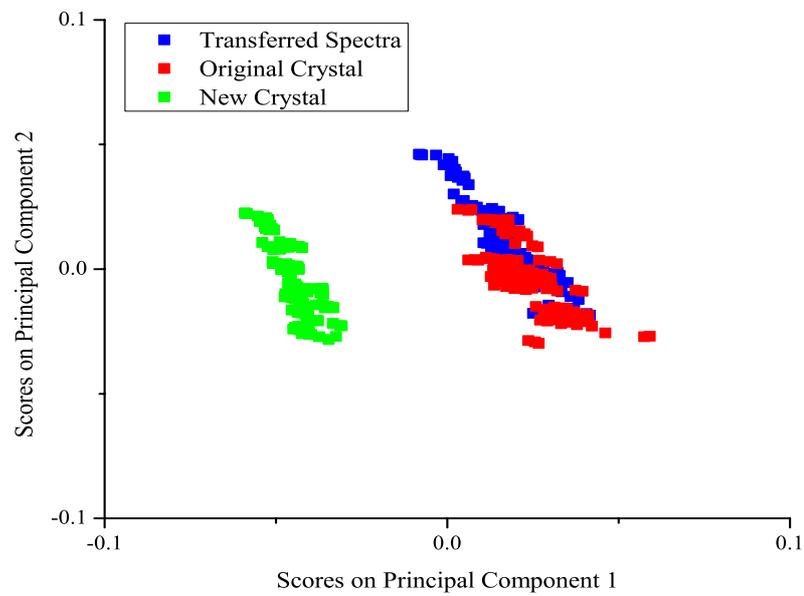
PDS, inverse sample selection original crystal to new crystal



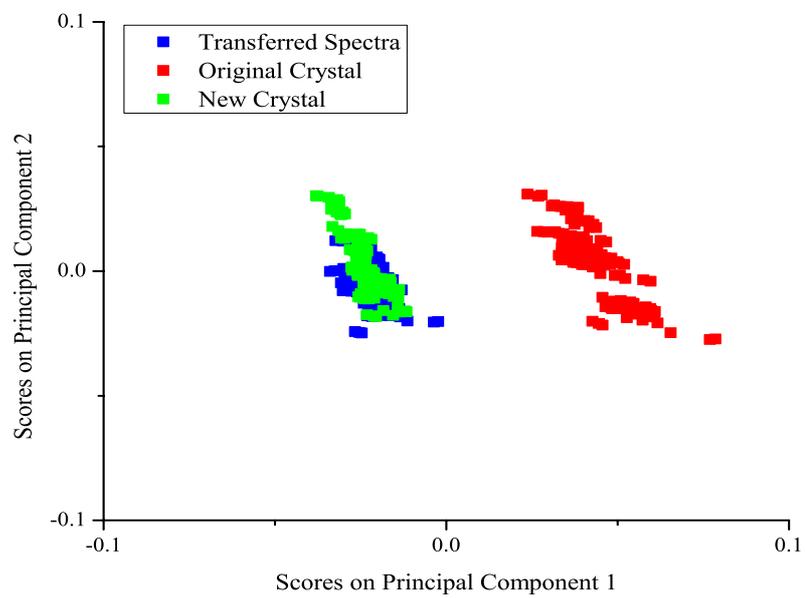
SST, leverage sample selection new crystal to original crystal.



STT, inverse sample selection new crystal to original crystal.



SST, inverse sample selection original crystal to new crystal.



Algor- ithm	Transfer Parameters	Transfer	Latent Variables	RMSEC (mMol dm ⁻³)			RMSECV (mMol dm ⁻³)			RMSEP (mMol dm ⁻³)		
				L- alanine	Cpd. A	Cpd. B	L- alanine	Cpd. A	Cpd. B	L- alanine	Cpd. A	Cpd. B
DS	Leverage	Old to New	5	39.9	57.8	32.4	41.2	60.4	33.7	85.9	69.8	24.4
DS	Leverage	New to Old	3	53.0	63.2	29.9	54.9	65.9	30.9	51.6	47.7	38.0
DS	Inverse	Old to New	5	55.6	109.4	40.2	59.6	115.7	41.7	69.1	45.1	33.7
DS	Inverse	New to Old	5	59.1	114.3	38.4	61.6	119.6	39.4	131.2	148.5	89.8
PDS	Leverage	Old to New	4	36.2	81.4	37.1	39.2	89.0	39.2	38.9	99.7	37.1
PDS	Leverage	New to Old	5	58.6	81.3	47.7	60.4	88.0	49.9	68.4	80.5	36.6
PDS	Inverse	Old to New	5	40.3	72.9	36.1	43.1	79.5	37.8	41.4	77.3	34.9
PDS	Inverse	New to Old	5	65.7	84.7	48.0	69.9	92.6	49.6	71.3	56.2	34.1
SST	Leverage	Old to New	6	56.4	92.6	29.6	60.8	99.5	30.6	65.6	86.5	41.6
SST	Leverage	New to Old	6	57.0	103.0	28.6	61.7	110.3	29.2	62.2	103.6	43.1
SST	Inverse	Old to New	6	53.8	86.5	27.5	57.4	94.1	28.3	65.4	124.9	40.5
SST	Inverse	New to Old	6	54.0	86.5	27.5	57.7	94.1	28.2	65.0	124.3	36.5

