

ULRR

A study of selected environmental issues related o biopharmaceutical manufacturing using Escherichia coli to produce a recombinant protein

Item Type	Thesis
Authors	Witt, Madlen K
Download date	2026-05-20 13:53:18
Item License	https://creativecommons.org/licenses/by-nc-sa/1.0/
Link to Item	https://hdl.handle.net/10344/4036



UNIVERSITY of LIMERICK

OLLSCOIL LUIMNIGH

**A study of selected environmental issues related
to biopharmaceutical manufacturing using
Escherichia coli to produce a recombinant
protein**

By

Madlen K. Witt B.Sc. (Hons)

Department of Chemical and Environmental Sciences

Supervisors:

Prof. Gary Walsh

Dr. Tom O'Dwyer

Thesis submitted to the University of Limerick in fulfilment of the
requirements of Doctoral of Philosophy

January 2014



Declaration

I hereby declare that his work is the result of my own investigations and that this report has not been submitted in this form or any other form to this or any other university in candidature for a higher degree.

January 2014

Acknowledgements

Firstly, I would like to express my sincerest thanks and gratitude to my supervisors, Prof. Gary Walsh and Dr. Tom O'Dwyer for their time, continuous guidance, expert help, support and encouragement throughout this project.

I greatly acknowledge the EPA, Ireland for funding this research. This work is funded as part of the Science, Technology, Research and Innovation for the Environment (STRIVE) Programme 2007-2013 which is financed by the Irish Government under the National Development Plan 2007-2013 and it is administered on behalf of the DEHLG by the Environmental Protection Agency (EPA) which has the statutory function of coordinating and promoting environmental research.

I would also like to thank my friends and colleagues in the University of Limerick Angela, Mike and Jayne for their help and friendship over the last number of years.

I wish to thank Daniel for his loyal friendship and understanding throughout this project and for always listening.

I wish to express a special gratitude to my family in particular my parents for their continued encouragement and love. Thank you so much for all your help, advice, patience and for always believing in me.

I would like to thank Marie for her support throughout all my studies as well as this project.

Finally, I want to thank Tadhg for his love, support and his endless patience. Thank you so much for being there, managing to put up with me and listening to my rants. I am so grateful to have you in my life.

Abstract

A study of selected environmental issues related to biopharmaceutical manufacturing using *Escherichia coli* to produce a recombinant protein

Author: Madlen Witt

Escherichia coli expression systems remain a preferred choice for the production of recombinant proteins for therapeutic, diagnostic and industrial purposes. Low costs and simplicity of culturing as well as straightforward genetic engineering technologies ensure their continued use for laboratory investigations as well as in commercial activities.

An *E. coli* expression system producing a recombinant protein was constructed for this research. The model strain, *E. coli* MC106 producing recombinant bacterial His₆-tagged β -galactosidase, was developed via standard genetic engineering techniques and protein expression was optimised to achieve high concentrations of soluble product.

Historically, during upstream processing little consideration was given to the potential environmental impacts of culture media ingredients which were often added in excess to achieve high cell density and hence product yields. The model *E. coli* strain was utilised to investigate the scope for reducing phosphorus (P) quantities included in a complex (LB and TB) and semi-defined (M9/YE) fermentation media. The findings showed that P reductions of up to 70 % did not adversely affect biomass and product yields attained; however, further P minimisation lead to a drop in dry cell weight as well as protein synthesis, particularly in the case of semi-defined media. Protein functionality, assessed by the kinetic parameters K_m and V_{max} , was not influenced by the type of media nor the P concentration present. 70 % P reductions would lead to significant P savings in large-scale manufacturing of proteins produced by genetically engineered *E. coli* strains.

The second part of this study entailed purification, at laboratory-scale to electrophoretic homogeneity, of the model protein via a traditional multi-chromatographic scheme and an affinity-based strategy. Both purification schemes were compared in terms of their environmental impact based on the buffers used. Utilising the engineered affinity-based approach reduced the number of downstream processing steps required to achieve purification from 6 to 4 and increased the final product yield from 11 % to 34 %. Environmental analysis of the chromatographic buffer constituents indicated that, per mg of purified protein, the use of the affinity-based method reduced the total P usage levels by 46 %, total ammonia by 99 %, total water usage by 75 % and total COD by 62 %, although the organic nitrogen levels increased by 75 %. In addition, comparative cost analysis showed a 60 % savings in chemical and chromatography costs per mg of purified product for this purification strategy. Although already widely used at research level, the use of affinity-based purification systems for process-scale protein purification would likely have significant environmental, energy and cost benefits. Furthermore, the study showed additional P savings can be achieved by using alternative buffering systems not containing P compounds during protein purification.

Mass balance simulations and environmental modelling was used to highlight the total phosphorus savings that can be achieved when employing the P-reduced fermentation media and the optimised purification strategies.

Table of Contents

Declaration	i
Acknowledgements	ii
Abstract	iii
Table of Contents	iv
List of Figures	ix
List of Tables	xiv
List of Abbreviations	xvii
List of Publications	xx
Chapter One – Introduction	1
1.1 General overview	2
1.2 Biotechnology sector in overview.....	2
1.3 Overview of biopharmaceutical production.....	5
1.4 Industrial recombinant protein expression in <i>Escherichia coli</i>	9
1.4.1 Upstream processing	14
1.4.1.1 Fermentation process of recombinant <i>E. coli</i>	14
1.4.1.2 Media requirements for <i>E. coli</i> expressing heterologous protein ..	17
1.4.1.3 Defined, semi-defined and complex media	20
1.4.2 Downstream processing	24
1.4.2.1 Chromatography	26
1.5 <i>E. coli</i> β -galactosidase.....	30
1.5.1 Structure of <i>E. coli</i> β -galactosidase	31
1.5.2 Mechanism of action.....	34
1.6 Environmental implications	36
1.6.1 Phosphorus in the environment.....	36
1.6.2 Phosphorus in the aquatic environment	39
1.6.2.1 Eutrophication.....	40
1.6.3 Phosphorus reserves	43
1.6.4 Integrated Pollution Prevention Control (IPPC)	44
1.7 Life cycle assessment (LCA)	45
1.7.1 Environmental evaluation models.....	47
1.8 Project aims and overview	48
Chapter Two – Generation of <i>E. coli</i> model strain	51
2.1 Introduction	52
2.2 Materials.....	53
2.2.1 Chemicals and reagents.....	53
2.2.2 Molecular biology reagents.....	53
2.2.3 Molecular strains	54
2.2.4 Media and supplements.....	55

2.3	Experimental methods	56
2.3.1	Microbial culturing techniques and cell lysis	56
2.3.2	Bioinformatic analysis	57
2.3.3	DNA isolation and sequencing	58
2.3.4	Polymerase chain reactions	59
2.3.5	Restriction enzyme digestion of DNA and ligation	60
2.3.6	Transformation	60
2.3.7	Optimisation of recombinant protein expression in <i>E. coli</i>	61
2.3.8	Polyacrylamide gel electrophoresis (PAGE)	62
2.3.8.1	Gel preparation and running	62
2.3.8.2	Native PAGE and zymogram	63
2.3.8.3	Western blotting	64
2.3.9	β -Galactosidase assay	64
2.3.10	Determination of temperature versus activity profiles	65
2.3.11	Determination of pH versus activity profiles	65
2.4	Results and Discussion	66
2.4.1	Genomic DNA purification	66
2.4.2	Amplification of <i>lacZ</i> gene using polymerase chain reaction	67
2.4.3	Purification and restriction of expression vector	68
2.4.4	Purification of PCR product and subsequent restriction	69
2.4.5	Ligation of PCR product with plasmid and transformation into <i>E. coli</i> competent cells	69
2.4.6	Sequencing analysis and homology modelling	71
2.4.7	Expression of recombinant β -galactosidase	72
2.4.8	Native PAGE and Western blot of recombinant β -galactosidase	76
2.4.9	Comparison of native and recombinant β -galactosidase	78
2.5	Conclusion	81
Chapter Three – Assessment of reducing the phosphorus content in fermentation media for recombinant protein production in <i>E. coli</i>		82
3.1	Introduction	83
3.2	Materials	84
3.3	Experimental methods	84
3.3.1	Fermentation and cell lysis	84
3.3.2	Absorbance values at 600 nm	84
3.3.3	Determination of dry cell weight (DCW)	84
3.3.4	Growth of <i>E. coli</i> on selected media	85
3.3.5	β -Galactosidase assay and kinetic analysis	85
3.3.6	Bradford assay	85
3.3.7	Determination of recombinant protein concentration from SDS-PAGE	86
3.3.8	Total phosphorus analysis	87
3.3.9	Statistical analysis	88
3.4	Results and Discussion	89

3.4.1	Media selection	89
3.4.2	Phosphorus analysis of all fermentation media.....	89
3.4.3	Effect of P minimisation in the fermentation media on <i>E. coli</i> biomass	92
3.4.3.1	Absorbance measurements.....	92
3.4.3.2	Growth profiles of model <i>E. coli</i> strain for all fermentation media	93
3.4.3.3	Dry cell weight.....	95
3.4.3.4	pH analysis of all media.....	97
3.4.4	Effect of P minimisation in the fermentation media on the recombinant protein	98
3.4.4.1	Concentration of the recombinant protein	98
3.4.4.2	Functional activity of recombinant protein.....	106
3.5	Conclusion	109
Chapter Four - A comparative environmental analysis of traditional multi-chromatographic versus affinity-based purification for a recombinant protein and reduction of the environmental impact of the affinity-based purification scheme		111
4.1	Introduction	112
4.2	Materials.....	113
4.3	Experimental methods.....	113
4.3.1	Fermentation and cell lysis	113
4.3.2	Traditional multi-chromatographic purification scheme	113
4.3.3	Affinity-based purification scheme.....	114
4.3.4	Chemical oxygen demand.....	115
4.3.5	Ammonia.....	116
4.3.6	Organic nitrogen	117
4.3.7	Minimisation of total phosphorus (TP) in waste streams produced during recombinant protein purification	117
4.4	Results and Discussion.....	118
4.4.1	Purification of recombinant β -galactosidase using a traditional multi-chromatographic scheme.....	118
4.4.2	Purification of recombinant β -galactosidase using an affinity-based scheme.....	124
4.4.3	Environmental analysis of both purification schemes	128
4.4.4	Cost analysis for both purification schemes	133
4.4.5	Reduction of environmental impact of affinity-based purification.....	134
4.5	Conclusion	142
Chapter Five – Mass balance simulation of phosphorus wastes arising from recombinant protein production using <i>E. coli</i>.....		144
5.1	Introduction	145
5.2	Materials.....	147

5.3	Methods.....	147
5.3.1	Mass balance.....	147
5.3.2	Upstream processing.....	147
5.3.3	Downstream processing.....	147
5.3.4	Environmental analyses of waste streams	147
5.3.5	SuperPro Designer [®] 6.0.....	148
5.4	Results and Discussion	149
5.4.1	Waste streams generated during the production of the recombinant model protein	149
5.4.1.1	Content of waste streams produced when employing the traditional multi-chromatographic purification scheme.....	149
5.4.1.2	Content of waste streams produced when employing affinity-based purification scheme.....	152
5.4.1.3	Content of waste streams produced when employing the P-reduced fermentation media and the affinity-based purification scheme with the alternative buffering system.....	155
5.4.2	Generation of mass balance for TP emitted throughout the recombinant protein production.....	157
5.4.2.1	Total phosphorus mass balance for recombinant protein production using traditional multi-chromatographic purification.....	158
5.4.2.2	Total phosphorus mass balance for recombinant protein production using affinity-based purification strategy	160
5.4.2.3	Total phosphorus mass balance for recombinant protein production using P-reduced fermentation media and the affinity-based purification scheme with the alternative buffering system.....	162
5.4.2.4	Total phosphorus in the waste streams of all production processes	164
5.4.3	Chemical oxygen demand of waste streams generated during recombinant protein production.....	166
5.5	Conclusions.....	169

Chapter Six – Environmental modelling of the phosphorus emissions from recombinant protein production processes using the EHS model 170

6.1	Introduction.....	171
6.2	Ecological hazard score (EHS) model structure	174
6.2.1	Goal and scope definition	174
6.2.2	Life cycle inventory analysis	176
6.2.3	Classification	176
6.2.4	Characterisation	177
6.2.5	Valuation.....	181
6.2.6	Ecological hazard score (EHS) calculation	185
6.3	Results and Discussions.....	188
6.3.1	Case study one: Application of the EHS model to the recombinant protein production using traditional multi-chromatographic and affinity-	

based purification particularly with respect to total phosphorus (TP) emissions.....	189
6.3.1.1 Goal and scope definition	189
6.3.1.2 Inventory analysis and impact assessment classification.....	192
6.3.1.3 Characterisation	194
6.3.1.4 Determination of hazard score (HS)	195
6.3.1.5 Valuation.....	197
6.3.1.6 Calculation of ecological hazard score (EHS).....	205
6.3.2 Case study two: Application of the EHS model to the production of recombinant protein production using affinity-based purification particular in terms of total phosphorus (TP) emissions	208
6.3.2.1 Comparison of P-reduced recombinant protein production process with standard process at laboratory-scale	213
6.4 Conclusion	215
Main findings and future work.....	217
Bibliography	221
Appendices.....	245
Appendix A: Protein and Gene sequences for β -galactosidase from <i>Escherichia coli</i> K12 substrain W3110 accession number AP_000996.1	246
Appendix B: Primer design for <i>lacZ</i> amplification from <i>E. coli</i> W3110.....	249
Appendix C: Restriction digest of <i>lacZ</i> from <i>E. coli</i> W3110 using NEBCutter	251
Appendix D: ClustalW alignment of recombinant <i>lacZ</i> -HT nucleotide sequence from MWG analysis and published gene sequence of <i>lacZ</i> from W3110 (Accession number AP_000996.1)	252
Appendix E: Amino acid alignment of β -galactosidase from <i>E. coli</i> W3110 (accession number AP_000996.1) and recombinant β -galactosidase.....	255

List of Figures

Figure 1.1 Overview of the production process of a biopharmaceutical product	13
Figure 1.2 Downstream processing during large-scale biopharmaceutical manufacturing detailing biological wastes arising	24
Figure 1.3 View of <i>E. coli</i> β -galactosidase tetramer identifying the twofold axes ...	32
Figure 1.4 Equation representing folding of <i>E. coli</i> β -galactosidase	33
Figure 1.5 General scheme of <i>E. coli</i> β -galactosidase action on the natural substrate lactose.....	35
Figure 1.6 The phosphorus cycle	38
Figure 1.7 Examples of soluble and insoluble phosphorus species found in natural waters	39
Figure 1.8 Life cycle assessment (LCA) framework	46
Figure 2.1 Map of pProEx-HTb plasmid	55
Figure 2.2 Electrophoresis analysis of purified genomic DNA from five overnight liquid cultures of <i>E. coli</i> W3110	66
Figure 2.3 Electrophoretic analysis of PCR amplifications using optimised conditions for <i>lacZ</i> gene.....	67
Figure 2.4 A 0.8 % (w/v) agarose gel depicting isolated pProEx-HTb and plasmid restriction results	68
Figure 2.5 Electrophoretic analysis of restricted amplicon using the restriction enzymes BamH1 and HindIII.....	69
Figure 2.6 Electrophoretic analysis of purified plasmid DNA from <i>E. coli</i> DH5 transformants.....	70
Figure 2.7 Purified and restricted plasmid DNA of <i>E. coli</i> DH5 transformants with pProEx-HTb carrying <i>lacZ</i> -HT insert visualised on 0.8 % (w/v) agarose gel.....	71
Figure 2.8 Excerpt of ClustalW alignment of recombinant <i>lacZ</i> -HT nucleotide sequence from the MWG analysis and published gene sequence of <i>lacZ</i> from W3110 (accession number AP_000996.1) showing the single base mutation (red)	72
Figure 2.9 SDS-PAGE analysis of recombinant β -galactosidase expressed in <i>E. coli</i> MC1061	73
Figure 2.10 Electrophoretic results of IPTG concentration optimisation studies for recombinant β -galactosidase expression.....	74
Figure 2.11 SDS-PAGE analysis of the crude extracts obtained during IPTG optimisation studies at 25 °C	75
Figure 2.12 Native gels showing crude extract from <i>E. coli</i> MC1061 pProEx-HTb_ <i>lacZ</i> -HT stained for protein and activity	76

Figure 2.13 SDS-PAGE and Western blot analysis of recombinant β -galactosidase expressed in <i>E. coli</i> MC1061	77
Figure 2.14 pH versus relative activity profile of the native and recombinant β -galactosidase enzyme	78
Figure 2.15 Temperature versus relative activity profile for the native and recombinant β -galactosidase enzyme.....	79
Figure 3.1 Growth curves of the model <i>E. coli</i> strain in complex fermentation media LB and TB 0-100 % P reduced	94
Figure 3.2 Growth curves of the model <i>E. coli</i> strain in semi-defined fermentation media M9/YE 0-100 % P reduced.....	95
Figure 3.3 SDS-PAGE analysis of standard β -galactosidase concentration for the production of the calibration curve	99
Figure 3.4 Standard calibration curve relating β -galactosidase loaded per lane on 100	
Figure 3.5 SDS-PAGE analysis of the soluble fractions for determination of the recombinant protein concentrations	102
Figure 3.6 SDS-PAGE gels of the non-soluble fractions of the model strain after fermentation during the P minimisation study	104
Figure 3.7 Electrophoretic analysis of the total protein fractions from the model strain obtained during the P minimisation study.....	106
Figure 4.1 Chromatogram for the purification of recombinant β -galactosidase using an IEC column containing DEAE Sepharose CL-6B.....	119
Figure 4.2 Chromatogram for the purification of recombinant β -galactosidase by HIC using Phenyl Sepharose HP.....	120
Figure 4.3 Chromatogram for the final step of the traditional multi-chromatographic scheme for recombinant β -galactosidase purification by SEC using Superdex200.	121
Figure 4.4 Electrophoretic analysis of recombinant β -galactosidase purified using traditional multi-chromatographic scheme.....	123
Figure 4.5 Chromatogram for the purification of recombinant β -galactosidase with IMAC containing Ni ²⁺ -charged Chelating Sepharose Fast Flow.....	125
Figure 4.6 Chromatogram for the final step of the affinity-based scheme for recombinant β -galactosidase purification by SEC using Superdex200	126
Figure 4.7 Electrophoretic analysis of recombinant β -galactosidase purified using the affinity-based chromatography scheme.....	127
Figure 4.8 Chromatogram showing the results of SEC runs for the recombinant β -galactosidase using buffers with varying phosphate salt concentrations	135
Figure 4.9 Chromatogram for the purification of recombinant β -galactosidase with IMAC using ACES salts as the alternative buffering system	138

Figure 4.10 Chromatogram for the purification of recombinant β -galactosidase with SEC using ACES salts as the alternative buffering system	139
Figure 4.11 Electrophoretic analysis of recombinant β -galactosidase purified using the affinity-based chromatography scheme with ACES buffering system	141
Figure 5.1 Flow schematic generated using SuperPro Designer [®] 6.0 highlighting the inputs and outputs of the recombinant β -galactosidase production process employing the traditional multi-chromatographic purification method.....	151
Figure 5.2 Flow schematic generated using SuperPro Designer [®] 6.0 highlighting the inputs and outputs of the recombinant β -galactosidase production process employing the affinity-based purification method.....	154
Figure 5.3 Flowsheet detailing TP mass balance for the recombinant protein production employing the traditional multi-chromatographic purification scheme	159
Figure 5.4 Flowsheet detailing TP mass balance for the recombinant protein production employing the affinity-based purification strategy.....	161
Figure 5.5 Flowsheet detailing TP mass balance for the recombinant protein production employing P-reduced TB media in combination with the affinity-based purification strategy and the alternative buffering system.....	163
Figure 6.1 Flowsheet depicting the steps involved in the ecological hazard score (EHS) model	173
Figure 6.2 The framework of the ecological hazard score (EHS) model.....	175
Figure 6.3 Schematic of the stages involved in the classification step for the determination of the ecological hazard score (EHS).....	176
Figure 6.4 Schematic overview of the stages involved in the characterisation step of the ecological hazard score (EHS) model	180
Figure 6.5 Steps of the analytical hierarchy process (AHP) application	181
Figure 6.6 Schematic outline of phase 1 of the valuation step for the determination of the ecological hazard score (EHS).....	182
Figure 6.7 Schematic overview of phase 2 of the valuation step for the determination of the ecological hazard score (EHS).....	183
Figure 6.8 Outline of phase 3 in the valuation step for the determination of the ecological hazard score (EHS).....	185
Figure 6.9 Outline of the calculation of the ecological hazard score (EHS).....	186
Figure 6.10 Procedure for computation of ecological hazard score (EHS)	187
Figure 6.11 Overview showing the application of the ecological hazard score (EHS) model during this study.....	188
Figure 6.12 Schematic boundaries (gate-to-gate) for recombinant β -galactosidase production using the traditional multi-chromatographic purification scheme.....	191

Figure 6.13 Schematic boundaries (gate-to-gate) for recombinant β -galactosidase production using the affinity-based purification strategy	192
Figure 6.14 Inventory and classification step for the recombinant β -galactosidase production process employing the traditional multi-chromatographic purification scheme analysed using EHS simulation model.....	193
Figure 6.15 Inventory and classification step for the recombinant β -galactosidase production process employing the affinity-based purification scheme analysed using EHS simulation model	194
Figure 6.16 Outline of characterisation step in the determination of EHS using simulation detailing specifically the computation of the equivalency factor (EF) for nutrient enrichment potential (NEP)	195
Figure 6.17 Outline showing hazard score (HS) calculation in the EHS simulation model for each of the waste streams generated during the recombinant protein production employing the traditional multi-chromatographic purification scheme.	196
Figure 6.18 Outline showing hazard score (HS) calculation in the EHS simulation model for each of the waste streams generated during the recombinant protein production employing the affinity-based purification scheme.....	197
Figure 6.19 Overview of valuation step in the determination of the EHS for the recombinant protein production using the traditional multi-chromatographic purification scheme	198
Figure 6.20 Overview of valuation step in the determination of the EHS for the recombinant protein production using the affinity-based purification scheme.....	199
Figure 6.21 The hierarchy structure for the emissions from the recombinant β -galactosidase production process using the traditional multi-chromatographic purification scheme	200
Figure 6.22 Overview of ecological hazard score (EHS) computation as determined for the recombinant protein production process employing the traditional multi-chromatographic purification scheme	206
Figure 6.23 Overview of ecological hazard score (EHS) computation as determined for the recombinant protein production process employing the affinity-based purification strategy.....	207
Figure 6.24 Inventory and classification step for the recombinant protein production process employing 70 % P-reduced TB media and the affinity-based purification scheme with the alternative buffering system analysed using EHS simulation model	209
Figure 6.25 Outline showing hazard score (HS) calculation in the EHS simulation model for each of the waste streams generated during the P-reduced recombinant protein production process	210

Figure 6.26 Overview of valuation step in the determination of the EHS for the recombinant protein production using affinity-based purification scheme with 70 % P-reduced TB media and the alternative buffers.....	211
Figure 6.27 Overview of ecological hazard score (EHS) computation as determined for the recombinant protein production process employing the affinity-based purification strategy with 70 % P-reduced TB media and the alternative buffering systems.....	212

List of Tables

Table 1.1 Types of expression systems used for recombinant protein production.....	7
Table 1.2 Nutrients required for <i>E. coli</i> growth in culture	18
Table 1.3 Chromatographic techniques commonly employed in protein purification	27
Table 2.1 Molecular biology reagents used during this study.....	53
Table 2.2 Microbial strains and plasmid used in this study.....	54
Table 2.3 Composition of media used during this study	56
Table 2.4 PCR primers designed to amplify <i>lacZ</i> from <i>E. coli</i> W3110	59
Table 2.5 PCR cycle parameters used for <i>lacZ</i> amplification.....	60
Table 3.1 Staining procedure for SDS-PAGE gels used for determination of recombinant protein concentration.....	86
Table 3.2 The chemical composition of the three media studied (LB, TB, M9/YE) along with the total phosphorus (TP) concentrations experimentally determined (section 3.3.8) in the P-containing media constituents.....	90
Table 3.3 The utilisation of total phosphorus (TP) by the model strain during growth and expression of the recombinant protein	91
Table 3.4 Total phosphorus (TP) concentrations experimentally determined for all media employed in this study.....	92
Table 3.5 Effect of P reduction in the fermentation media for the model <i>E. coli</i> strain upon optical density values (OD ₆₀₀) observed after recombinant protein expression (6 h post induction).....	93
Table 3.6 Reduction of P and its effects on mean dry cell weight (DCW) values obtained after fermentation using the various media	96
Table 3.7 Mean pH values of all media employed during the study pre and post fermentation	97
Table 3.8 Reduction of P in the fermentation media and its effects on soluble recombinant protein concentrations	101
Table 3.9 Reduction of P in the fermentation media and its effects on non-soluble recombinant protein concentrations	103
Table 3.10 Reduction of P in the fermentation media and its effects on total recombinant protein concentrations	105
Table 3.11 β -Galactosidase activity values of the soluble crude fraction obtained during the P minimisation study.....	107
Table 3.12 Effect of P reduction on recombinant protein's kinetic parameters K_m and V_{max}	108

Table 4.1 Purification table for recombinant β -galactosidase purified using the traditional multi-chromatographic scheme	122
Table 4.2 Purification table for recombinant β -galactosidase purified using the affinity-based chromatography scheme	126
Table 4.3 Environmental analysis of all buffers employed during the purification of the recombinant β -galactosidase	129
Table 4.4 Environmental analysis of the various buffers employed and comparison of the two purification schemes	131
Table 4.5 Cost breakdown of the two purification schemes for the isolation of the recombinant β -galactosidase	134
Table 4.6 Summary of results obtained during minimisation of P salts in SEC buffer	136
Table 4.7 Environmental analyses of alternative buffering systems (50 mM concentration) for the use of recombinant protein purification via the affinity-based scheme.....	137
Table 4.8 Purification table for recombinant β -galactosidase purified using the affinity-based chromatography scheme with ACES buffers.....	140
Table 5.1 Summary of the waste constituents present in the aqueous wastes 1-6 (W_1 - W_6) and in the solid waste (SW) generated during the recombinant β -galactosidase production using the traditional multi-chromatographic purification scheme.....	150
Table 5.2 Summary of the waste constituents present in the aqueous wastes 1-4 (W_1 - W_4) and in the solid waste (SW) generated during the recombinant β -galactosidase production using the affinity-based purification scheme	153
Table 5.3 Summary of the waste constituents present in the aqueous wastes 1-4 (W_1 - W_4) and in the solid waste (SW) generated during the recombinant β -galactosidase production using the P-reduced fermentation media and the affinity-based purification scheme with the alternative buffering system	156
Table 5.4 Total phosphorus (TP) analyses of the aqueous waste streams generated from all three laboratory-scale recombinant protein production processes	165
Table 5.5 Chemical oxygen demand (COD) analyses of the aqueous waste streams generated from all three laboratory-scale recombinant protein production processes	167
Table 6.1 The category quantitative assessment of the ecological impact indices .	179
Table 6.2 The standardised numerical comparison scale for the construction of pair-wise comparison matrices	184
Table 6.3 Pair-wise comparison matrix for ecological impact categories (level 2) with respect to the goal (level 1)	201

Table 6.4 Ecological hazard score (EHS) for the four waste streams (W₁-W₄) investigated for the standard and P-reduced recombinant protein production process with respect to total phosphorus emissions 213

List of Abbreviations

2xYT	Double yeast extract and tryptone media
ACES	N-(2-Acetamido)-2-aminoethanesulfonic acid
ADA	N-(2-Acetamido) iminodiacetic acid
AF	Affinity chromatography
AHP	Analytical hierarchy process
ANOVA	Analysis of variance
AP	Acidification potential
API	Active pharmaceutical ingredient
APS	Ammonium persulphate
ATP	Adenosine triphosphate
ATR	Acid tolerance response
BAT	Best available technique
BES	N, N-Bis(2-hydroxyethyl)-2-aminoethanesulfonic acid
BHK	Baby hamster kidney
Bis-Tris	2, 2-Bis(hydroxymethyl)-2, 2, 2-nitrilotriethanol
Bis-Tris propane	2, 2'-(Propane-1, 3-diyldiimino)bis(2-(hydroxymethyl)propane-1, 3-diol)
BLAST	Basic local alignment search tool
bp	Base pairs
BPEO	Best practical environmental option
BSA	Bovine serum albumin
C	Carbon
CHO	Chinese hamster ovary
COD	Chemical oxygen demand
CR	Consistency ratio
C-terminal	Carboxyl-terminal
CV	Column volume
DCW	Dry cell weight
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen (German collection of microorganisms and cell cultures)
DSP	Downstream processing
DTT	Dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EBPR	Enhanced biological phosphorus removal
EC	European Commission
EEA	European Environment Agency
EF	Equivalency factor
EFTA	European Free Trade Association
EHS	Ecological hazard score
ELV	Emission limit value
EMA	European Medicines Agency
EMB	Eosin methylene blue
EMBL	European Molecular Biology Laboratory
EMS	Environmental management system
E-PRTR	European Pollutant Release and Transfer Register
EU	European Union

FDA	U. S. Food and Drug Administration
FRP	Filterable reactive phosphorus
g(DCW) l ⁻¹	Gram of dry cell weight per litre
gDNA	Genomic deoxyribonucleic acid
GRAS	Generally regarded as safe
GST	Glutathione S-transferase
GWP	Global warming potential
HCDC	High cell density culture
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIC	Hydrophobic interaction chromatography
His ₆ -tag	Polyhistidine tag
HS	Hazard score
IEC	Ion exchange chromatography
IMAC	Immobilised metal affinity chromatography
IPPC	Integrated pollution prevention control
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ISO	International Organisation for Standardisation
kb	Kilo base pairs
K _m	Michaelis constant
<i>lacZ</i>	Gene for β-galactosidase enzyme
<i>lacZ</i> -HT	Gene for β-galactosidase enzyme with polyhistidine tag
LB	Luria-Bertani broth
LCA	Life cycle assessment
LPS	Lipopolysaccharide
M9	Minimal media 9
M9/YE	Minimal media 9 supplemented with yeast extract
MES	2-(N-morpholino)ethanesulfonic acid
mg(β-gal) g(DCW) ⁻¹	Milligrams of soluble β-galactosidase per gram of dry cell weight
MM	Molecular mass
N	Nitrogen
NADH	Nicotinamide adenine dinucleotide
NCBI	National Centre for Biotechnology Information
NEP	Nutrient enrichment potential
NH ₃	Ammonia
N _{org}	Organic nitrogen
N-terminal	Amino-terminal
O	Oxygen
OD ₆₀₀	Optical Density at 600 nm
ODP	Ozone depletion potential
OECD	Organisation for Economic Co-operation and Development
ONP	ortho-nitrophenol
ONPG	ortho-nitrophenol-β-D-galactosidase
P	Phosphorus
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PE	Phase effect factor
PF	Prioritisation factor
pI	Isoelectric point

PIPES	1, 4-Piperazinediethanesulfonic acid
POCP	Photochemical ozone creation potential
PTM	Post-translational modifications
psi	Pounds per square inch
<i>p</i> -value	Probability value
PVDF	Polyvinylidene fluoride
Q	Quantity
RE	Restriction enzyme
RNA	Ribonucleic acid
RP	Rock phosphate
rpm	Rounds per minute
SAIC	Science Applications International Corporation
SB	Superbroth
SCBD	Secretariat of the Convention on Biological Diversity
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate- Polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SW	Solid waste
TB	Terrific broth
TCA	Tricarboxylic acid cycle
TEMED	Tetramethylethylenediamine
TFP	Total filterable phosphorus
TIM	Triosephosphate isomerise
TMB	3, 3', 5, 5'-Tetramethylbenzidine
TP	Total phosphorus
Tricine	N-(2-Hydroxy-1, 1-bis(hydroxymethyl)ethyl)glycine
Tris	2-Amino-2-hydroxymethyl-propane-1, 3-diol
TRP	Total reactive phosphorus
TSS	Total suspended solids
TY	Tryptone and yeast extract media
UF	Ultrafiltration
UF/DF	Ultrafiltration in combination with diafiltration
UV	Ultraviolet
v/v	Volume per volume
V_{max}	Maximum velocity
w/v	Weight per volume
W_1	Waste stream one
W_2	Waste stream two
W_3	Waste stream three
W_4	Waste stream four
W_5	Waste stream five
W_6	Waste stream six
WFI	Water for injection
WWTP	Wastewater treatment plant
X-gal	5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside
YE	Yeast extract
β -gal	β -galactosidase

List of Publications

Submitted/In preparation

Witt, M., O'Dwyer, T. and Walsh, G., 'Minimisation of phosphorus in the fermentation media of *Escherichia coli* producing a model recombinant protein' in preparation for *International Journal of Environmental Science and Technology* (submission number JEST-D-12-01092)

Witt, M., O'Dwyer, T. and Walsh, G., 'Environmental comparison of two downstream strategies for purification of a recombinant protein from *Escherichia coli*' in preparation for *Journal of Cleaner Production*

Meeting abstracts

Witt, M., O'Dwyer, T. and Walsh, G. (2012), 'Minimisation of phosphorus in the fermentation media of recombinant *Escherichia coli* producing a protein', The 22nd Irish Environmental Researchers' Colloquium (Environ 2012), University College Dublin, 7th-9th March 2012, p. 92 (Oral Presentation)

Witt, M., O'Dwyer, T. and Walsh, G. (2012), 'Reduction of phosphorus waste produced during microbial upstream processing', Proceedings of the 5th International Conference on Sustainable Energy and Environmental Protection (SEEP 2012), Part-II, Dublin City University, 5th-8th June 2012, p. 265 (Poster Presentation)

Witt, M., O'Dwyer, T. and Walsh, G. (2010), 'Generation of a model system to allow for an environmental analysis of biopharmaceutical manufacturing using recombinant *Escherichia coli*', The 20th Irish Environmental Researchers' Colloquium (Environ 2010), Limerick Institute of Technology, 17th-19th February 2010, p. 116 (Poster Presentation)

Chapter One – Introduction

1.1 General overview

Escherichia coli (*E. coli*) expression systems are regularly used in the biotechnology industry to produce recombinant proteins, e.g. biopharmaceuticals. The biopharmaceuticals sector makes up one of the most important sections in the biotechnology industry and is estimated to reach global market value of around \$166 billion by 2017 [IMARC 2012]. A variety of medical conditions are treated using these therapeutic proteins including cancer, diabetes, growth disturbances, haemophilia and hepatitis [Walsh 2009; Walsh 2010a].

Industrial-scale manufacturing of *E. coli*-produced proteins entails the propagation of a strain, capable of producing the recombinant product, in a fermentation bioreactor. This is followed by harvesting of the cells and subsequent purification of the final product. Throughout this process biological and chemical impurities/wastes are generated and removed using a variety of methods. The biological impurities are similar for all recombinant proteins produced and include cellular debris, nucleic acids, contaminating proteins and lipopolysaccharides. Regulatory authorities state that these impurities must be reduced and/or removed for the final material to be fit for clinical use [EMA 1999; FDA 2013a]. The chemical wastes that are generated during the fermentation, purification and post production steps (equipment cleaning, decontamination and sanitisation) depend on the specific protein/process. Waste streams that arise from these processes require deactivation or capturing of any microorganism prior to release into the environment, for example, liquid wastes are usually inactivated using heat treatment. This is followed by wastewater treatment where compounds containing nutrients such as phosphorus (P) can be removed and possibly recovered. The release of P into water bodies can have detrimental effects to the surrounding environment with eutrophication being the main concern. As these subjects are broad and diverse, this introduction will focus on the topics which are relevant to this specific research project.

1.2 Biotechnology sector in overview

Biotechnology is defined as any 'technological application that uses biological systems, living organisms or derivatives thereof, to make or modify products or

processes for specific use' [SCBD 2013]. This industrial sector has gained significant interest since the discovery of hybridoma technology and genetic engineering in the mid 1970s as this allowed the integration of traditional industrial microbiology, i.e. production of fermented materials and purification of native products and intermediated metabolites, with molecular biology [Demain 2007].

The use of recombinant DNA technology and protein engineering enables the production of proteins which specifically suit the user or process requirements. This greatly impacted on the enzyme and agricultural industry as well as medicine with the production of products in the fields of healthcare, diagnostics, food, nutrition, detergents, textiles, leather, paper, pulp, polymer and plastics [Demain and Vaishnav 2009]. These products include primary metabolites from microbes such as amino acids, nucleic acids, vitamins, organic acids, alcohols and polysaccharides as well as secondary metabolites, e.g. enzyme inhibitors, immunosuppressants, anticancer agents, bioherbicides and bioinsecticides [Demain 2007].

The involvement of biotechnology in so many areas encouraged a considerable growth of the industry and resulted in increased investment activities towards new companies dedicated to genetic approaches and in the establishment of hundreds of start-up biotechnology companies by the early 1980s [Demain 2000; Walsh 2000]. The global biotechnology market reached a value of \$281.7 billion by 2011 and is forecasted to increase by more than 60 % by 2016 [MarketLine 2012].

The biggest subsector within biotechnology is the production of proteins for commercial application via recombinant means. This can be broken down into three categories: bulk industrial proteins (mainly enzymes), analytical/diagnostic proteins and proteins for therapeutic purposes [Walsh 2002; Soetaert and Vandamme 2010].

The global market for proteins with industrial applications was valued at \$2.5 billion in 2009 and is believed to reach \$3.7 billion by the end of 2013 [Demain and Vaishnav 2009]. These enzymes have applications in detergents, pulp and paper, textiles, leather, agriculture and feeds, chemicals and food industries [Soetaert and Vandamme 2010]. The market leading catalytic proteins are proteases which accounts for 57 % of market shares. Other enzymes include amylase, glucoamylase, xylose isomerase, lactase, lipase, cellulase, pullulanase and xylanase. Industrial

proteins are produced in large quantities in the magnitude of several thousand tons annually [Demain and Vaishnav 2009]. Some small amounts are still produced in animals and plants, however, genetic manipulation allows manufacturing of desired qualities and quantities of these enzymes in microbes. Therefore, most of the enzymes are produced by GRAS (generally regarded as safe)-status microorganisms through submerged or solid-state fermentation. This type of cultivation is simpler, faster and often economically favourable which is of critical importance for their commercial success. Industrial enzymes are only partially purified at best and sold as concentrated liquids or granulated dry products [Demain and Vaishnav 2009; Soetaert and Vandamme 2010].

Proteins produced for analytical/diagnostic applications include hundreds of enzymes and antibodies produced for assay systems. These are important for analysis of clinical samples such as blood, serum and urine for the detection and monitoring of diseases. However, enzymatic test kits are also designed for industrial and scientific environments with a multitude of purposes for sample or product analysis. Additionally, this group of biotechnology products includes proteins employed during molecular techniques such as restriction enzymes and polymerases [Kopetzki *et al.* 1994; Demain 2007]. Diagnostic/analytical proteins are manufactured at much lower quantities than industrial enzymes but need to be highly purified due to the nature of their application [Soetaert and Vandamme 2010].

Therapeutic proteins, i.e. biopharmaceuticals, make up the largest segment of the biotechnology market, accounting for 67.4 % of the total market value [MarketLine 2012]. Biopharmaceuticals are defined as any 'biology-based therapeutic that structurally mimics compounds found within the body' [Nagle *et al.* 2008]. This includes all recombinant therapeutic proteins, monoclonal antibody-based products used for *in vivo* medical purposes and nucleic acid-based medicinal products [Walsh 2003b]. Throughout the last century, a large number of biomolecules were shown to have therapeutic potential. However, most proteins which could be used in medical treatments are present naturally in low quantities and chemical synthesis or semi-synthesis is technically challenging and/or economically unsuitable. Recombinant DNA technology overcame those difficulties and enabled large-scale production while facilitating the development of purposefully modified/re-engineered proteins

which can display clinical advantages over the native protein product [OECD 1998; Walsh 2003b; Han *et al.* 2010; Kamionka 2011]. Over 210 biopharmaceutical products have gained approval for medical use in Europe and/or United States and typically 8-14 new biopharmaceuticals are approved by the global regulatory authorities each year [Walsh 2010b; Walsh 2012; Rader 2013]. Drug approval trends most likely reflect commercial rather than technical considerations and target indications over the last 10 years include diabetes, hepatitis, cancer prevention and treatment, abnormal growth-related conditions and genetic defects [Walsh 2006; Nagle *et al.* 2008; Walsh 2012]. Polypeptide drug disadvantages include low bioavailability, thus requiring injection for many available drugs to date, and high cost, but their advantages of high specificity and low toxicity far outweigh their negative aspects [Müller *et al.* 1996; Demain 2007].

1.3 Overview of biopharmaceutical production

The process development for any biopharmaceutical is initiated with the target molecule discovery [Josic *et al.* 2008; Walsh 2012]. The “target” molecule is closely associated with the disease, as it is either a missing, an overproduced or malfunctioning protein or its derivative. A therapeutic hypothesis is formulated and screening techniques are employed to identify a viable drug candidate [Gill 2009]. In recent years combined transcriptome and proteome approaches have played a major role in gaining useful information during target identification [Han *et al.* 2010]. The human genome project also impacted on the biomedical research and its implication for healthcare. Molecular medicine has shown the potential to improve disease diagnosis and can help with the creation and selection of targeted treatments. Furthermore, techniques such as DNA microarray, bioinformatics and chemogenomics are used in the aid of drug design and development [Wu-Pong and Rojanasakul 2008].

The next step is at the research and development stage with the construction of an expression cell line capable of producing the selected target protein. Using genetic engineering, the production cell system is designed often via the introduction of a plasmid housing the nucleotide sequence that expresses the protein of interest or chromosomal integration of the gene of interest into the cell genome [Walsh 2003b;

Dowd *et al.* 2007]. A major component at this stage is the choice of expression system. There are a number of different expression systems capable of producing recombinant biopharmaceutical products. These are listed in Table 1.1 together with their advantages and disadvantages. In general, the primary aspects for choosing one over the other are determined by the functional activity requirements of the protein product. For therapeutic proteins often post-translational modification (PTM) requirements, in particular glycosylation, need to be considered [Walsh 2003b]. Glycosylation, i.e. the covalent addition of a carbohydrate component, is species-, tissue- and cell-type specific and can affect a number of functionally significant characteristics. These include correct protein folding and assembly, targeting and trafficking of the protein to its final destination, ligand binding and stability of the protein, making it of vital importance for a therapeutic product. All PTMs lead to a distinct biological outcome and other commonly observed PTMs in proteins are: acetylation, acylation, ADP-ribosylation, amidation, γ -carboxyglutamate formation, disulphide bond formation, hydroxylation, phosphorylation and sulfation [Soetaert and Vandamme 2010; Walsh 2010b]. Additional factors important for choosing an expression system include economics, available expertise and infrastructure and protein features, such as source, size, solubility and refolding behaviour [Ho and Gibaldi 2004]. New advances in proteome analysis provide understanding regarding changes in protein synthesis, degradation rates, post-translational modifications and protein interactions which all help during expression cell line design [Han *et al.* 2010]. Recombinant protein expression in *E. coli* is further discussed in section 1.4 of this introduction.

Once the recombinant production system has been constructed, the large-scale manufacturing procedure for the biopharmaceutical can be developed. Production processes need to be scalable, reproducible and relatively inexpensive with appropriate manufacturing conditions. The safety and functionality of the product is ensured by including stringent quality control and quality assurance measures which are highly regulated and rigorously controlled. Industrial-scale production of any biopharmaceutical involves the product biosynthesis phase (upstream processing) which is followed by the purification phase and generation of the finished product (downstream processing) [Walsh 2003b].

Table 1.1 Types of expression systems used for recombinant protein production

Expression system	Examples	Advantages	Disadvantages
Prokaryotic cell cultures	<i>Escherichia coli</i>	<ul style="list-style-type: none">• Well characterised systems that allow quick and precise modifications to the genome• Promoter control is relatively straightforward• Appropriate fermentation technology available• High levels of expression possible (accumulation of heterologous protein up to 50 % of dry cell weight)• Cultures grow rapidly on inexpensive media to high cell densities• Survival is possible under variety of environmental conditions• Protease activity is easily reduced	<ul style="list-style-type: none">• Inability to carry out post-translational modifications (PTM)• Presence of lipopolysaccharides (LPS)• High level expression may lead to insoluble aggregates (inclusion bodies)
Mammalian cell cultures	Chinese hamster ovary (CHO) and baby hamster kidney (BHK)	<ul style="list-style-type: none">• Native structure, including PTM, therefore, usually best for biological activity• Wide range of mammalian expression vectors	<ul style="list-style-type: none">• Possible low yields• High production costs due to complex nutritional requirements and slow growth• Susceptible to physical cell damage• Added risk of contaminating product with animal pathogens
Yeast cell cultures	<i>Saccharomyces cerevisiae</i> and <i>Pichia pastoris</i>	<ul style="list-style-type: none">• Classification as GRAS (generally regarded as safe)-listed organisms and no detectable lipopolysaccharides• Grow relatively quickly on inexpensive media• Studied in detail (genetic manipulation techniques and fermentation equipment available)• Stable integration of expression cassette into the host genome (multiple copies possible)• Only 0.5 % of endogenous proteins secreted, which simplifies downstream processing• Ability to carry out PTM	<ul style="list-style-type: none">• Glycosylation pattern often differs from native human glycoproteins• Expression levels can be relatively low
Fungi cell cultures	<i>Aspergillus</i> strains	<ul style="list-style-type: none">• Suitable fermentation technology available• High levels of expression possible• Allow excretion into extracellular media (advantage during downstream processing)• Ability to carry out PTM	<ul style="list-style-type: none">• Glycosylation pattern often differs from native human glycoproteins• Produce extracellular proteases which may degrade the recombinant protein

Expression system	Examples	Advantages	Disadvantages
Insect cell cultures	Fall Armyworm (Spodoptera frugiperda), Drosophila melanogaster (fruitfly)	<ul style="list-style-type: none"> • High levels of intracellular recombinant protein (due to powerful viral promoters) with recombinant protein expression levels of 30-50 % of total intracellular protein content • Expression vectors designed to attack invertebrates only (ensuring safety) • Human pathogens generally do not infect insect cell lines • Use relatively inexpensive media • Can be cultured rapidly 	<ul style="list-style-type: none"> • Low expression of extracellular levels (which would simplify purification) • PTM often differs from native human glycoproteins or can be incomplete
Transgenic animals	Goat, rabbit, sheep and pig	<ul style="list-style-type: none"> • Ease of harvesting of crude product • Low capital investment and running costs • Potential for high expression levels • Ongoing supply guaranteed 	<ul style="list-style-type: none"> • Often great variability in expression levels • Significant lag time between transgenic embryo and commencement of routine product manufacture
Transgenic plants	Tobacco, potato and rice	<ul style="list-style-type: none"> • Low costs for plant cultivation • Well established and inexpensive harvesting methodologies and equipment • Ease of scale-up • Proteins expressed in seeds are stable • Free of human pathogens 	<ul style="list-style-type: none"> • Low or variable expression levels • Possibility of post-translational gene silencing • Glycosylation pattern differs from native human glycoproteins • Possible seasonal/geographical plant growth issues • Potential presence of plant metabolites that could contaminate the crude product • Environmental/public concern relating to potential environmental escape of genetically altered plants
Cell-free (<i>in vitro</i>)	<i>E. coli</i> lysate, wheat germ lysate, reticulocyte lysates of rabbits	<ul style="list-style-type: none"> • Faster than <i>in vivo</i> expression systems • PTM possible depending on cell extracts • Enable protein labelling with modified amino acids • Expression of proteins that undergo rapid proteolytic degradation by intracellular proteases or toxic proteins • Simple to express many different proteins simultaneously • Use of additives such as metal ions, detergents, cofactors, binding partners 	<ul style="list-style-type: none"> • Low yield • Possible depletion of substrates and accumulation of inhibitory products limits reaction time • Novel technology which can be constrained to small reaction volumes and therefore render it non-suitable for large-scale production • Relatively high price-performance ratio

Adapted from Walsh [2007b], Soetaert and Vandamme [2010] and Subramanian [2012]

1.4 Industrial recombinant protein expression in *Escherichia coli*

Escherichia coli is the most studied prokaryotic organism and widely used for genetic manipulations and production of proteins with therapeutic and commercial applications. This expression host is the first choice for laboratory investigations and in the initial development of commercial activities or just as a comparison for various expression platforms [Baneyx 2004; Chen 2012]. Production of therapeutic proteins in *E. coli* expression systems displays a number of technical advantages (Table 1.1) and is therefore commonly used. As of 2012, 213 individual biopharmaceuticals have gained regulatory approval in Europe and the United States and of these 63 (29.6 %) are produced in *E. coli* [Ryan, M. personal communication]. Quantitatively, this translates to approximately two thirds of biopharmaceutical APIs (active pharmaceutical ingredients) are produced in microbial systems, mainly *E. coli* [Ransohoff 2010]. The genomes of various *E. coli* strains have been sequenced and the information is readily available in public databases. Another advantage of the this expression system is the large number of well-characterised expression vectors available [Pissarra 2004; Kolaj *et al.* 2009]. Additionally, many therapeutic proteins, including insulin, human growth hormones and a number of cytokines, are already produced in this bacterium and approval of *E. coli*-based products by the U.S. Food and Drug Administration (FDA) or European Medicines Agency (EMA) is relatively routine [Baneyx 2004].

The position of this organism in the biopharmaceutical industry was strengthened when advances in the understanding of protein folding and translocation allowed secretion of soluble proteins into the periplasm where the oxidising environment catalyses the formation of disulphide bonds (most biopharmaceuticals naturally contain disulphide linkages). Furthermore, the periplasmic space contains fewer proteases, many of which target specific substrates, providing a more stable environment for the heterologous product. Also protein purification of the recombinant product expressed in the periplasm is simpler as there is a lower protein fraction present. If cytoplasmic expression is required then molecular manipulation allows blocking of the reducing pathways in the cytoplasm additionally to the co-overexpression of proteins which exhibit protein disulphide isomerase activity. This enables the formation of stable disulphide bonds in the cytoplasm in these mutants

[Lee 1996; Baneyx 2004; Kolaj *et al.* 2009]. Co-production of molecular chaperones and folding catalyst often improves expression of heterologous proteins in *E. coli* and co-production systems which include six to seven folding modulators are used to overcome folding bottlenecks [Kolaj *et al.* 2009].

One of the key considerations for choosing *E. coli* as host cells is their high growth rate of 3.33 cell division cycles per hour under optimal conditions compared to yeast or mammalian cells which have a growth rate of 0.25 and 0.02 cell division cycles per hour, respectively [Ho and Gibaldi 2004]. This allows *E. coli* to grow rapidly in bioreactors to high cell densities on relatively inexpensive liquid substrates in a relatively short time by standard methods and to produce high levels of heterologous proteins [Baneyx 1999; Walsh 2003b].

With *E. coli* expression systems using plasmids is the preferred means, rather than chromosomal integration, and a large number of vectors have been constructed and optimised. The use of plasmids ensures that high gene copy numbers within each cell can be achieved which in turn improves product yield. The exact copy number depends on the plasmids, for example, ColE1-derived plasmids are usually present at 15-30 copies while the pUC series show 500-700 copies per cell. For rapid downstream processing, expression vectors can produce fusion proteins with a wide range of affinity tags which support protein purification. Fusion partners can also allow detection of the product or improve the solubility of their passenger products that may otherwise accumulate as inclusion bodies in the cell cytoplasm [Baneyx 1999; Baneyx 2004; Hunt 2005]. The loss of plasmids during large-scale fermentation is a major industrial problem as it limits/reduces product yields. Plasmid instability is caused by the unequal distribution of the vectors to daughter cells during growth and cell division. Cells without a plasmid construct have a smaller energy burden allowing them to grow faster and eventually dominate the culture. Stable maintenance of the plasmids in cells is guaranteed through the addition of antibiotics in the fermentation media which is inactivated by a vector encoding resistance gene. However, this adds economic costs and disposal of the spent media proves a potential environmental hazard as the antibiotic and the antibiotic resistance gene may be released. That problem is avoided when an

essential gene of the chromosomal DNA of the host bacterium is deleted while at the same time this gene is added to the plasmid [Baneyx 2004; Glick *et al.* 2010].

There are various disadvantages to the *E. coli* expression system including its inability to undertake post-translational modifications most notably glycosylation [Pissarra 2004]. As previously mentioned, glycosylation can play an important part in the protein's biological activity. However, some glycoproteins do not require the glycan component to function (e. g. recombinant interleukin-2) and in these cases *E. coli* expression systems are used. Other strategies for expression of recombinant glycoproteins in *E. coli* involves mutation of residues on the surface to create more soluble proteins (preventing aggregation due to non-glycosylation) or mutation of glycosylation sites to cysteines which facilitates subsequent *in vitro* glycosylation [Kamionka 2011]. Additionally, work is being carried out to incorporate bacterial glycosylation systems into *E. coli* for glycosylation of non-native proteins [Chen 2012]. Another major problem is the formation of inclusion bodies in the cytoplasmic or periplasmic space when overexpressing heterologous proteins in *E. coli*. These inclusion bodies are insoluble, aggregated and inactive forms of the product, usually processing non-native intra- and intermolecular disulphide bridges and unusually free cysteines [Soetaert and Vandamme 2010]. One of the attempts made to reduce the occurrence of inclusion bodies involves reducing the temperature of growth and/or expression from 37 °C to 30 °C or lower. This decreases the rate of biochemical reactions, including transcription and translation, providing time for correct protein folding. Another method entails the fusion of the recombinant product to the *E. coli* thioredoxin gene which eliminates the inclusion body formation in some instances [Jana and Deb 2005; Walsh 2007b]. Expression of the recombinant protein into the periplasm, as described earlier, may also increase the concentration of soluble functional product.

Of further concern with *E. coli*-derived products is the possible contamination of recombinant protein preparations with lipopolysaccharides (LPS or endotoxin). LPS molecules can cause pyogenic and shock reactions in humans if not removed from the biopharmaceutical product. However, mutant *E. coli* strains which exhibit a reduced presence of LPS have been designed and downstream processing procedures

which facilitate the removal of these contaminants are available [Baneyx 2004; Walsh 2007b].

In addition, documented during protein production in *E. coli* is membrane leakage and cell lysis which leads to significantly reduced yields. Another disadvantage for *E. coli* expression systems is the extensive patent portfolio which may require negotiation with multiple licensing agreements [Baneyx 2004; Kolaj *et al.* 2009].

Despite those drawbacks many of the top selling biopharmaceuticals, such as insulin, human growth factors and interferons, as well as new therapeutic proteins are produced in *E. coli* due to its many technical and economical advantages [Walsh 2010a]. *E. coli*-based systems remain the default choice for the production of biopharmaceuticals that do not require post-translational modifications. As a result, a great amount of technical data has been obtained [Walsh 2003a] including ways to increase product yield with high cell density cultivation and strategies for fed-batch microbial fermentation [Knorre *et al.* 1991]. The improved understanding of function, regulation and interactions of cellular gene products and protein folding pathways, as well as the new genetic tools, ensure that *E. coli* is one of the most utilised expression hosts for heterologous proteins in the biopharmaceutical industry [Baneyx 2004].

Once the recombinant *E. coli* expression system has been generated, process development continues with the determination of optimal production conditions and protein isolation options focussing on product yield. To meet pharmaceutical needs the host cells are cultured large-scale up to kilograms and sometimes kilotons depending on the medical application [Ho and Gibaldi 2004]. The development of a production strategy starts at laboratory- and moves to pilot plant-scale (100-19,000 l scale). During these stages parameters such as temperature, pH, rate and nature of mixing and oxygen demand in the bioreactor are established. Optimal conditions generally change with each 10-fold increase in volume and therefore require extensive testing with each scale-up. There are also technical considerations to address such as bioreactor design, which needs to ensure adequate sterility as well as inclusion of probes for continuous online monitoring [Glick *et al.* 2010]. After all conditions for a scalable and reproducible expression are finalised, scale-up to production level (1,000-100,000 l) commences [Junker 2004].

The manufacturing process of a biopharmaceutical product can be divided into two distinct steps: ‘upstream’ and ‘downstream’ processing [Soetaert and Vandamme 2010]. In *E. coli* expression systems, upstream processing refers to the initial propagation of the strain which results in the formation of the protein product in a fermentation vessel. The downstream phase involves recovery and purification of the protein and generation of the finished product [Walsh 2003b]. The stages of biopharmaceutical production in *E. coli* are illustrated in Figure 1.1.

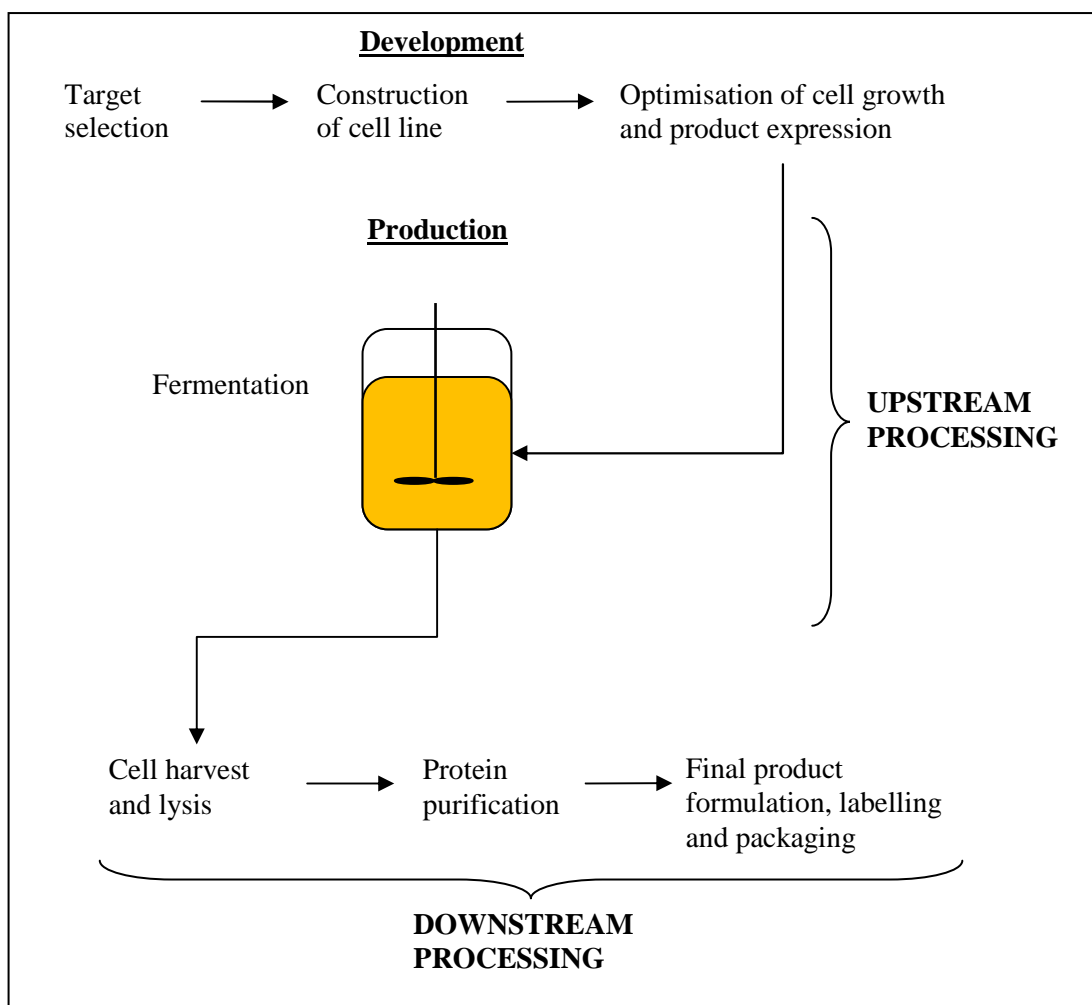


Figure 1.1 Overview of the production process of a biopharmaceutical product
Adapted from Walsh [2007b]

1.4.1 Upstream processing

The upstream processing, also called “live” area, begins with the inoculation of sterile media with the producer organism followed by incubation under optimum conditions. After sufficient growth this pre-culture will be used to inoculate the “seed” fermentor which is subsequently utilised to cultivate the production bioreactor [Walsh 2003b]. Industrial bioreactors can be quite large, therefore, the inoculum needs to be built-up through several successive stages to 5-10 % of the working volume of the production fermentor. Using this approach ensures optimal use of the production bioreactor by minimising the batch time in this vessel [Soetaert and Vandamme 2010]. The cell growth and target production conditions would have been established during the initial product development, typically through optimisation studies [Dowd *et al.* 2007]. Characteristic production-scale bioreactors for recombinant DNA products range from 10,000 litres upwards [Junker 2004].

1.4.1.1 Fermentation process of recombinant *E. coli*

E. coli can be cultured using one of three types of fermentation processes: batch, fed-batch or continuous. In batch operations a specific volume of culture is inoculated with the microbial culture and the fermentation proceeds for a certain time, referred to as the batch or fermentation time [Soetaert and Vandamme 2010]. This type of operation is characterised by no addition to or withdrawal from the culture with the exception of the gas phase and often acid/base solutions for pH control [Cinar 2003]. Under these conditions six typical phases of microbial growth can be observed: lag phase (1), acceleration phase (2), logarithmic/exponential phase (3), deceleration phase (4), stationary phase (5) and death phase (6) [Glick *et al.* 2010]. In the lag phase the microorganism adapts to the new environment and little or no growth is observed. This phase can be time consuming and therefore in large-scale production costly. To minimise this stage the inoculum is often grown in a comparable medium, a large amount of inoculum is used to keep the dilution shock small and the cells of the inoculum are in the late logarithmic stage [Macauley-Patrick and Finn 2008; Soetaert and Vandamme 2010]. This is followed by an active growth period (phases 2-4) during which the cell mass undergoes several doublings. At this stage primary metabolites essential for microbial survival are produced. After that the culture enters the stationary phase either due to depletion of a critical growth substance or due to

accumulation of metabolic end products that inhibit cell growth. During this phase the biomass remains constant, however, the cellular metabolism changes and secondary metabolites are synthesised. At the end of this stage the cells energy reserves are depleted and the culture enters the death phase [Cinar 2003; Glick *et al.* 2010]. Batch fermentation generally operates over four to five days for biopharmaceutical production [Soetaert and Vandamme 2010] and is typically terminated at the end of the growth phases or during the stationary phase depending on when the product is formed and overall product yields [Cinar 2003]. This type of fermentation is the simplest mode of operation and allows a substantial quantity of cell biomass and product to be obtained. It is historically established and most industrial-scale fermentation is carried out using batch operations. However, there are some disadvantages to batch fermentation and these include batch-to-batch variability and limitations to initial substrate concentrations due to inhibition and repression effects of some nutrients. Another problem is necessary downtime required on account of cleaning, sterilisation, filling and cooling of equipment as it increases production costs [Macauley-Patrick and Finn 2008; Glick *et al.* 2010]. Some therapeutic proteins produced in laboratory environments using batch fermentation include human epidermal growth factor, human interleukin-7, protective antigen protein and human necrosis factor- [Choi *et al.* 2006].

In fed-batch fermentation, sterile culture medium, specific nutrients and/or inducers are added continuously or periodically to the production bioreactor without any culture withdrawal [Cinar 2003; Soetaert and Vandamme 2010]. Employing this feeding strategy allows the control of the specific growth rate and formation of unwanted by-products. This in turn prolongs the logarithmic and stationary phase resulting in increased biomass and amount of metabolites, including product, formed [Macauley-Patrick and Finn 2008]. Also, some nutrients required for high cell biomass can repress or inhibit cell growth if present at too high concentrations [Lee 1996]. Fed-batch operation allows the slow addition of these nutrients. Some microorganisms produce proteolytic enzymes during the stationary phase that can degrade the recombinant protein, therefore, it is important that the fermentation reaction for these recombinant microorganisms does not reach this part of the growth phase. This mode of fermentation requires more monitoring and control compared to batch operations. Another concern during this large-scale cultivation is the unequal

distribution of the high concentration fed solution which may result in gradients of essential variables (substrate concentration, dissolved oxygen and pH) [Soini *et al.* 2008]. Nevertheless, due to the increased yields that can potentially be obtained it is becoming more popular [Glick *et al.* 2010]. Fed-batch fermentation technologies were employed for the production of the following therapeutic proteins at laboratory-scale: bone morphogenetic protein 2, human interferon- α and - β and human tissue type plasminogen activator [Choi *et al.* 2006].

In a continuous culture operation, sterile medium or specific nutrients essential for growth are constantly added and a portion of the culture is continuously withdrawn for product recovery [Cinar 2003]. In this steady-state condition the total volume in the bioreactor as well as the number of cells present remain constant because the cells lost due to product removal are replaced by new cells through cell division using the fresh nutrients [Glick *et al.* 2010]. This fermentation is typically preceded by batch cultures and feedings start once the biomass has reached a certain level [Soetaert and Vandamme 2010]. Continuous operation prolongs the logarithmic phase, i.e. production formation period, while maintaining an environment with fewer fluctuations in nutrient concentration and overall biomass. This results in higher product yields and consistency compared to batch operations [El-Mansi *et al.* 2007; Macauley-Patrick and Finn 2008]. Employing this mode of fermentation also has economical benefits as it eliminates the inherent downtime associated with batch and fed-batch operations and utilises smaller bioreactors and purification equipment [El-Mansi *et al.* 2007; Glick *et al.* 2010]. However, maintaining sterile conditions during large-scale production for long periods can be difficult and contamination is often a problem [El-Mansi *et al.* 2007]. Some laboratory-scale continuous cultures were employed to produce therapeutic proteins, e.g. human interferon- α [Srivastava and Mukherjee 2005].

In addition to the mode of fermentation operation, the establishment of high cell density cultivation (HCDC) are typical strategies during large-scale recombinant protein production. HCDC is the method of choice for commercial heterologous protein production in *E. coli*, due to the high volumetric productivity that is associated with this method. Volumetric productivity depends on the cell biomass as well as specific productivity, i.e. the amount of product formed per unit cell per unit

time. Therefore, increasing the cellular density will increase the overall volumetric productivity of the process [Shojaosadati *et al.* 2008]. HCDC techniques for growing *E. coli* are well established and facilitate the efficient, high level production of various proteins. Cell concentrations of more than 50 g of dry cell weight per litre ($\text{g(DCW)} \text{ l}^{-1}$) are routinely achieved and methods are in place to obtain biomass concentrations of $150 \text{ g(DCW)} \text{ l}^{-1}$ [Glick *et al.* 2010]. Advantages of HCDC include increased cost effectiveness, reduced culture volume, which in turn ensures easier downstream processing, reduced investment in equipment and a decrease in the amount of waste produced. However, the HCDC methods have drawbacks such as substrate inhibition or limitation, limited oxygen transfer capacity and high oxygen demand, formation of growth inhibitory by-products, limited heat dissipation, plasmid instability and cell lysis and proteolysis. Despite those disadvantages, the economical benefits of HCDC are often greater and steps are taken to overcome these problems [Lee 1996; Shojaosadati *et al.* 2008]. That includes online monitoring and tight control of fermentation variables such as pH, temperature, dissolved oxygen and some critical nutrient concentrations as well as optimisation of media composition and nutrient feeding strategy, the use of different promoters to regulate levels of expression or use of different host strains [Walsh 2003b; Choi *et al.* 2006].

1.4.1.2 Media requirements for *E. coli* expressing heterologous protein

The media composition for recombinant *E. coli* cultivation can considerably influence the amount of biomass produced and thereby the heterologous protein yield [Nikerel *et al.* 2006]. The nutrient constituents affect metabolic pathway fluxes in the bacteria which not only impacts on maximum attainable cell yield but also the specific productivity for recombinant proteins and formation of by-products [Choi *et al.* 2006]. Therefore, optimisation of nutrient requirements for recombinant *E. coli* expressing the protein of interest is vital to achieve HCDC during large-scale production. Other factors in media determination are availability and quality consistency of ingredients, economics and the nature of the product, i.e. is the product intra- or extracellular [Walsh 2003b; Soetaert and Vandamme 2010].

E. coli is a non-fastidious microorganism that grows in both rich complex organic media as well as salt-based chemically defined media as long as a source of organic carbon is provided [Nikerel *et al.* 2006]. The general chemical formula for microbial

cells is $CH_xO_yN_z$ where x, y and z variables range from 1.65-1.85, 0.12-0.25 and 0.26-0.56, respectively. Thus, to support growth and protein production the fermentation media must contain carbon (C), oxygen (O) and nitrogen (N) [Soetaert and Vandamme 2010]. The full nutritional requirements for *E. coli* are well-defined and documented [Prescott *et al.* 2002; Madigan *et al.* 2009] and presented in Table 1.2.

Table 1.2 Nutrients required for *E. coli* growth in culture

Element	Chemical form supplied in culture media	Functions
Carbon (C)	Glucose, amino acids, complex mixtures	Main constituent in cellular material
Hydrogen (H)	H ₂ O, organic compounds	Constituent of organic compounds and cell water
Oxygen (O)	H ₂ O, O ₂ , organic compounds	Constituent of cell material, O ₂ is an electron acceptor in aerobic respiration
Nitrogen (N)	Inorganic: NH ₄ Cl, (NH ₄) ₂ SO ₄ Organic: amino acids, N bases of nucleotides	Constituent of amino acids, nucleic acids, some carbohydrates and lipids, enzyme cofactors and other substances
Phosphorus (P)	KH ₂ PO ₄ , K ₂ HPO ₄ , Na ₂ HPO ₄	Constituent in nucleic acids, phospholipids, nucleotides, enzymes cofactors, some proteins and other cell components
Sulphur (S)	Na ₂ SO ₄ , Na ₂ S ₂ O ₃ , Na ₂ S, organic sulphur compounds	Constituents in amino acids cysteine and methionine, some carbohydrates and coenzymes
Potassium (K)	KCl, KH ₂ PO ₄ , K ₂ HPO ₄	Cellular inorganic cation and cofactor for certain enzymes
Magnesium (Mg)	MgCl ₂ , MgSO ₄	Cellular inorganic cation and cofactor for certain enzymes
Iron (Fe)	FeCl ₃ , FeSO ₄	Key component in cytochromes and cofactor for enzymes and electron-carrying proteins
Micronutrients	Cobalt, Copper, Manganese, Molybdenum, Nickel, Zinc, etc.	Cofactors for enzymatic reactions

Adapted from Prescott et al. [2002] and Madigan et al. [2009]

The macronutrients C, O and N together with phosphorus (P) and sulphur (S) are required as components of carbohydrates, lipids, proteins and nucleic acids, i.e. cellular building blocks [Prescott *et al.* 2002; Madigan *et al.* 2009]. Carbon is supplied in form of sugars such as glucose or sucrose, starch products, glycerol, molasses or yeast extract. Approximately 50 % of the C source is incorporated into the cell biomass while the remaining C is used to derive energy for biosynthesis. Oxygen and hydrogen are also derived from the carbon sources as well as from water and through additional aeration [Heinzle *et al.* 2007]. Nitrogen is delivered to fermentation cultures through inorganic salts or organic compounds containing nitrogen [Davies and Demain 1999; Soetaert and Vandamme 2010]. However, ammonia is the preferred form as it can be readily converted to usable forms by enzymes like glutamine synthetase and glutamate synthase. Phosphorus is required for the formation of nucleotides for nucleic acids, phospholipids, cofactors and some proteins. Inorganic phosphorus can be directly incorporated by the bacteria and thus is often supplied in this form. Potassium (K), magnesium (Mg), calcium (Ca), sodium (Na) and iron (Fe) are also essential as they exist as cations and fulfil various functions within the cell. Additionally, trace amounts of micronutrients such as cobalt (Co), copper (Cu), manganese (Mn), molybdenum (Mo), nickel (Ni) and zinc (Zn) are necessary as they act as cofactors in enzymatic reactions. These are often provided through inorganic salts or are part of organic complex components [Prescott *et al.* 2002; Madigan *et al.* 2009].

High concentrations of these nutrients are needed to accommodate the nutritional requirements of large-scale cultures. However, most nutrients can have adverse effects on cell growth and production if present above a certain concentration [Lee 1996]. Precipitation of media ingredients (e.g. Mg and Fe) can also be a problem especially when present at high concentrations as it hinders nutrient supply, may interfere with monitoring devices and can affect downstream processing [Shiloach and Fass 2005]. High ionic concentration in growth media may result in osmotic pressures and conductivity that can affect the membrane potential and activate different stress mechanisms that induce a decrease in growth rate or termination of the growth cycle. It is therefore necessary to design a balanced medium with all nutrients below the inhibitory threshold or develop a feeding strategy that avoids precipitation and facilitates HCDC [Shojaosadati *et al.* 2008].

E. coli is generally grown under aerobic conditions since anaerobic growth provides less energy for metabolic processes such as protein synthesis and generates metabolites which further impair cell growth and result in reduced recombinant product yields [Tripathi *et al.* 2009b; Krause *et al.* 2010]. The expression of more than 200 genes by *E. coli* is reliant on oxygen availability [Losen *et al.* 2004] and to ensure oxygen does not become the rate limiting factor in bioreactors, the dissolved oxygen levels are monitored and aeration is supplied constantly [Tripathi *et al.* 2009b]. Due to the high oxygen demand in HCDC systems, oxygen is often provided through oxygen-enriched air or pure oxygen [Choi *et al.* 2006].

An additional problem during HCDC fermentation is the formation of acetate which is known as the bacterial “Crabtree” effect [Luli and Strohl 1990; Shiloach and Fass 2005]. *E. coli* produces acetic acid as an extracellular co-product of aerobic fermentation, which exists as the acetate ion at neutral pH [Eiteman and Altman 2006]. Acetate is produced when the carbon flux into the central metabolic pathway exceeds the biosynthetic demands and the capacity for energy generation within the cell [Lee 1996]. The bacteria will use the production of acetate from acetyl Coenzyme A to meet its energy requirements as this pathway generates large amounts of ATP and NADH [Lischke *et al.* 1993]. The main reason for acetate metabolism may be the overloading of the tricarboxylic acid (TCA) cycle by fast oxidation through glycolysis and/or saturation of the electron transport process [Shiloach and Fass 2005]. The accumulation of acetate may lead to a severe drop in pH and was correlated to reduced recombinant protein production [Macauley-Patrick and Finn 2008]. Acetate formation can be reduced by controlling the oxygen levels during the fermentation or by replacing glucose with fructose, glycerol or other carbon sources [Aristidou *et al.* 1999; Demain and Vaishnav 2009]. Other techniques in reducing acetate accumulation involve using pressurized culture regimes, which increase the oxygen transfer, or the use of genetic engineering to design mutants which produce less acetate or even consume acetate as a carbon source [Shojaosadati *et al.* 2008].

1.4.1.3 Defined, semi-defined and complex media

Media employed during *E. coli* fermentation can be one of three types depending on the nature of the materials used: synthetic/chemically defined media, semi-defined

and complex media [Zhang and Greasham 1999]. Chemically defined media consists of inorganic salt-compounds, trace metals and a carbon source. In these media the exact chemical compositions are known and can be controlled to obtain high cell densities. Complex media contains ingredients of imprecise or varying chemical composition like yeast extract and peptones which makes this media less reproducible [Lee 1996; Prescott *et al.* 2002]. Yeast extract is the water-soluble portion of autolysed yeast and peptone an enzymatic digest of proteins [Madigan *et al.* 2009]. Both components serve as sources of carbon, energy and nitrogen and may also contain growth factors and vitamins [Prescott *et al.* 2002; Krause *et al.* 2010]. Semi-defined media consists mostly of defined components with only one or two complex ingredients [Zhang and Greasham 1999].

Complex media such as Luria-Bertani broth (LB) and Terrific broth (TB) are well established for *E. coli* fermentation and protein expression [Sambrook and Russell 2001]. The media are easy to prepare and the ingredients are inexpensive and readily available [Krause *et al.* 2010; Soetaert and Vandamme 2010]. Terrific broth, devised by Tartoff and Hobbs [1987], was specifically developed to achieve high-level expression. It provides higher nutrient content and improved buffering which supports higher cell densities and a longer period of logarithmic growth phase for *E. coli* when compared to other media (e. g. LB) [Losen *et al.* 2004; Hunt 2005; Fu *et al.* 2006; Tripathi *et al.* 2008; Kensy *et al.* 2009]. Other rich media designed to achieve high cell densities are Super Broth (SB), tryptone-yeast extract (TY) medium and double yeast extract and tryptone (2xYT) media. Their high concentration in complex nutrients and the addition of phosphate supports high cell densities [Ausubel *et al.* 1988; Sambrook and Russell 2001; Tripathi *et al.* 2008; Tripathi *et al.* 2009b]. Bacterial cultures grow much faster on these media as their complex ingredients provide key intermediates and building blocks [Kensy *et al.* 2009]. However, lot-to-lot variability associated with these ill-defined components results in variations during fermentation and/or requires extensive “use” testing of the raw materials [Zhang *et al.* 2003; Shojaosadati *et al.* 2008].

Chemically defined media are often employed for recombinant bacterial cultivation, although they are known to support slower cell growth and produce lower protein yields [Shojaosadati *et al.* 2008]. Nonetheless, they are common practise as they can

generate high cell densities with more consistent titres, allow easier control and monitoring and often simplify downstream processing [Tripathi *et al.* 2009a]. There are many different chemically defined media that can be used for the cultivation and maintenance of *E. coli* [Atlas 2004]. All include a carbon source with various concentrations of inorganic salts depending on the strain and purpose of study [Lee 1996; Aristidou *et al.* 1999; Sambrook and Russell 2001]. The addition of single or few amino acids as well as vitamins may be necessary to improve cell and recombinant protein product yields [Shojaosadati *et al.* 2008]. One of the most commonly utilised chemically defined media or a variation thereof is minimal media 9 (M9) [Sambrook and Russell 2001; Tripathi *et al.* 2009b].

Semi-defined media is often employed for *E. coli* fermentation as the addition of complex nutrients (such as yeast extract or tryptone) increases biomass yield and boosts product formation [Shojaosadati *et al.* 2008]. Yeast extract allows acetate assimilation and growth even after glucose depletion and supplementation of media with peptones can stabilise enzymatic activity [Nancib *et al.* 1991].

The exact media composition or fermentation process for industrial-scale production is not readily available in either company or regulatory literature [EMA 2013; FDA 2013b] as it is considered confidential in commercial manufacturing. However, the media is similar to small-scale cultivation, i.e. consists of a combination of carbon, nitrogen, phosphorus, sulphur, minerals and essential vitamins. Carbon is usually present at a concentration range of 0.5-20 % (w/v), nitrogen 0.1-10 % (w/v), phosphorus 0.1-2 % (w/v), sulphur 0.1-1 % (w/v) and other nutrient are present at less than 1 % (w/v) [Davies and Demain 1999]. Large-scale fermentation media may also include chelating agents such as EDTA to prevent precipitation of ions, antifoam agents and inducers which initiate product formation [Waites *et al.* 2001; Heinzle *et al.* 2007]. The initial step in media optimisation is based on the stoichiometry of the *E. coli* growth and product formation. This includes an evaluation of the inputs, i.e. carbon, nitrogen, oxygen and minerals, and their conversion to cell biomass, metabolic products, carbon dioxide and heat. From that, it is possible to determine the minimum amount of nutrients required for a specific quantity of biomass and recombinant protein [Waites *et al.* 2001]. Subsequently, basic approaches such as trial-and-error processes or statistical techniques are used to optimise the media

composition further for maximum biomass yield and recombinant protein expression [Nikerel *et al.* 2006; Shojaosadati *et al.* 2008]. The preference of individual raw materials employed for industrial manufacturing depends on a number of factors [Waites *et al.* 2001]:

- Cost and availability (materials should be inexpensive, supplied at consistent quality and available all year around),
- Ease of handling in solid/liquid stage, transport and storage,
- Sterilisation requirements and possible denaturation concerns,
- Formulation, mixing, complexing and viscosity characteristics that could influence agitation, aeration and foaming during fermentation and downstream processing,
- Concentration of product protein attained, i.e. its rate of formation and yield per gram of substrate utilised,
- Levels and range of impurities present (potential generation of undesired by-products during the process),
- Overall health and safety implications,
- Regulatory considerations which for example encourage the use of defined media and media free of animal-derived components.

Pure chemicals can be used during small-scale expression but these provide a large economic burden during industrial manufacturing [Waites *et al.* 2001]. Complex ingredients are less expensive and have been used in many industrial fermentations [Davies and Demain 1999; Soetaert and Vandamme 2010], however, their undefined nature can increase costs and complexity during product recovery and purification [Waites *et al.* 2001].

In conclusion, it is likely that for biopharmaceuticals expressed in recombinant *E. coli* strains both complex and chemically defined media are utilised. However, the scientific literature suggests a move towards chemically defined media, which is also more acceptable from a regulatory and quality control standpoint.

1.4.2 Downstream processing

Downstream processing (DSP) follows upstream fermentation and comprises of recovery of the heterologous protein, concentration and final product formulation [Walsh 2007b]. The complexity of DSP depends on the nature of the product and its intended application. Recombinant therapeutic proteins expressed in *E. coli* are present at dilute concentrations in complex mixtures and require extremely high purity (> 95 %) which translates into many consecutive processing steps [Forciniti 2008]. Therefore, the cost of recovery and purification of these products is high and can involve as much as 50-80 % of total manufacturing costs, especially if the protein is produced intracellularly [Lowe *et al.* 2001; Shuler and Kargi 2002]. The individual steps in these multistage operations are closely related to the type and nature of the target protein and are considered highly confidential by the manufacturer [Walsh 2003b]. Figure 1.2 outlines the general stages involved in isolation and purification of biopharmaceutical products.

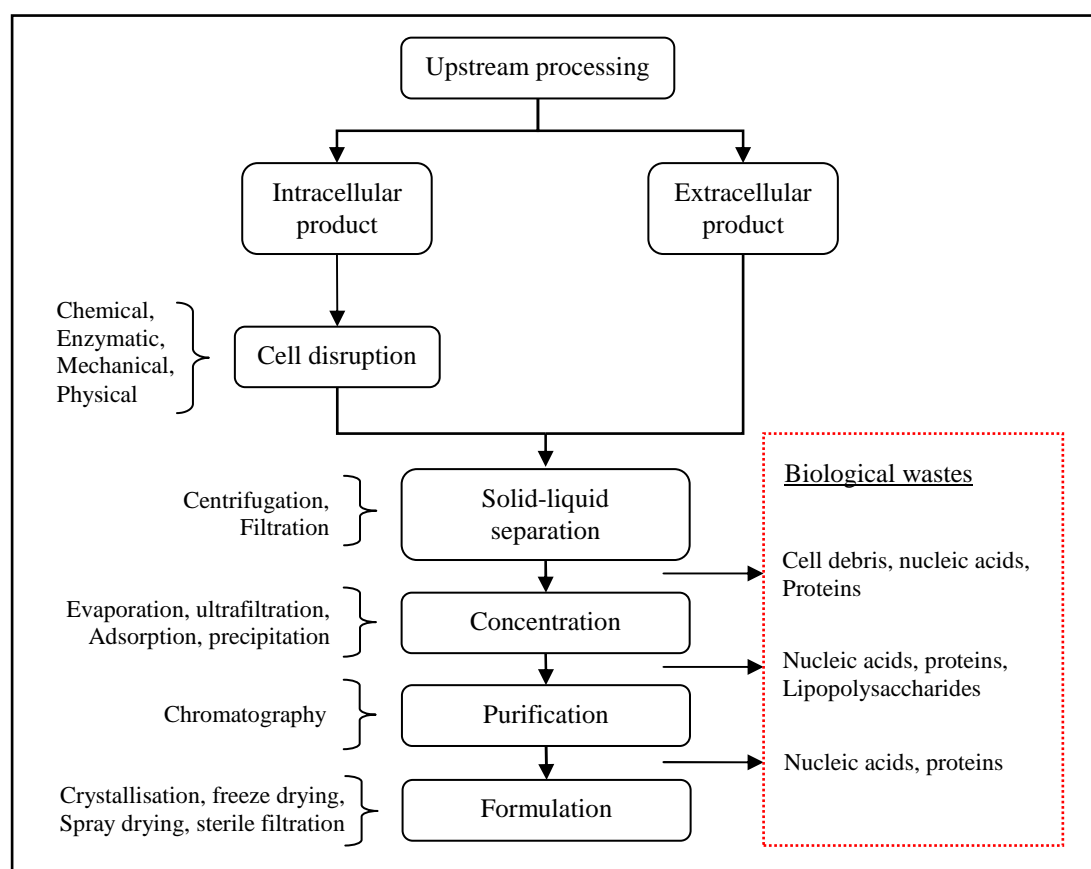


Figure 1.2 Downstream processing during large-scale biopharmaceutical manufacturing detailing the arising biological wastes

Adapted from Ratledge and Kristiansen [2001]

DSP is initiated by the separation of the soluble product from the cell biomass through solid-liquid separation. This process is carried out using centrifugation, filtration or more recently microfiltration and removes the microbial mass from the fermentation culture. However, most recombinant proteins are retained within the cell and cell wall disruption is required prior to the removal of the cellular debris [Forciniti 2008; Soetaert and Vandamme 2010]. There are many different techniques available for cell wall disruption at laboratory-scale using mechanical and non-mechanical methods [Flickinger and Drew 1999]. However, cell disintegration in industrial environments is carried out using chemical (detergent), enzymatic (lysozyme) or in the majority of cases, due to costs, mechanical (bead mill and high-pressure homogenizer) methods [Hubbuch and Kula 2007; Forciniti 2008].

Process economics dictate water removal at this stage, i.e. concentration of the crude protein, as it will be more convenient to work with, allowing greater speed and minimisation of the equipment for the following steps [Shuler and Kargı 2002; Walsh 2007b]. Concentration may be achieved via precipitation, using salts or solvents, adsorption techniques or most commonly employed ultrafiltration [Walsh 2007b].

The target protein is then purified using a combination of extraction, precipitation, a variation of chromatographic and/or ultrafiltration steps [Walsh 2003b; Hubbuch and Kula 2007]. Liquid chromatography employed during DSP of biopharmaceutical products is further discussed in section 1.4.2.1. During these steps contaminants such as host cell proteins, cell wall residues (including LPS), nucleic acids, viruses and truncated, clipped or chemically modified forms of the target protein are removed [Hubbuch and Kula 2007]. The absolute purity required for therapeutic proteins depends on their intended use, the dose, the risk-benefit ratio, etc. Optimised DSP schemes achieve purities of 98-99 % for recombinant proteins [Walsh 2007b; Carta and Jungbauer 2010]. These high purity demands together with technical issues such as equipment size and/or requirements of buffer and other raw material volumes can, therefore, contribute extensively to the manufacturing costs [Hubbuch and Kula 2007]. Decreasing the number of steps necessary for successful purification will beneficially impact on the economical as well as environmental concerns of large-scale recombinant protein production.

The concluding stages of DSP consist of the potential addition of excipients, filtration for product sterilisation and final formulation for storage (e.g. lyophilisation, crystallisation) [Walsh 2007b].

1.4.2.1 Chromatography

Chromatography is the principle unit operation used during the capture, purification and polishing steps observed in current industrial processes of biopharmaceutical manufacturing [Carta and Jungbauer 2010]. In column chromatography the components are separated based on their differential distribution between two phases: the solid stationary phase (chromatographic media/resin) and the liquid mobile phase (usually a chromatographic solvent/buffer) [Walsh 2007b]. This technique provides high resolution efficiencies which permits separation of complex mixtures with similar molecular properties. Chromatographic media consists of high capacity absorbents capable of capturing target proteins from low concentration solutions regularly experienced in DSP. Additionally, chromatography can be carried out using closed systems and the chromatographic resin can be regenerated. The methods are well established in biopharmaceutical manufacturing and suitable equipment and materials are readily available [Carta and Jungbauer 2010]. Advances in chromatographic materials and technical understanding often overcome difficulties encountered during scale-up and ensure this unit operation's continued use [Cramer and Holstein 2011].

Biopharmaceutical DSP typically requires a combination of two to four different chromatographic steps to achieve the necessary purity [Walsh 2007b]. There is an extensive collection of chromatographic media available that allow separation based on almost all physicochemical properties of proteins (see Table 1.3) [Walter 1999]. During the optimisation of target protein purification various chromatographic resins and methods are tested for binding and elution efficiency as well as removal of contaminants [Edwards 1997]. The chromatographic techniques employed in this study are further discussed below.

Size exclusion chromatography (SEC or gel filtration) separates proteins in solution based on their three-dimensional shape and size, i.e. their hydrodynamic volume [Forciniti 2008]. The resin consists of spherical particles with pores of varying sizes

and differential permeation through these pores is the basis of gel filtration [Davies and Demain 1999]. It is considered the simplest and mildest chromatography technique as there is no adsorption of the molecules to the medium [GE Healthcare 2010a]. The mobile phase is a fixed composition buffer and can be varied to suit the sample requirements or demands for further DSP stages [Flickinger and Drew 1999]. SEC facilitates resolution of aggregated forms of proteins making it a useful polishing step during purification. However, this chromatographic technique has a relatively low sample capacity and column packing as well as sample loading can be more demanding than other methods. Additional concerns are the significant dilution of the protein sample at the end of elution and the low flow rates that need to be used resulting in long processing times [Davies and Demain 1999; Walsh 2007b; GE Healthcare 2010a].

Table 1.3 Chromatographic techniques commonly employed in protein purification

Technique	Separation Principle
Ion exchange chromatography	Differences in net surface charge
Size exclusion chromatography	Differences in hydrodynamic volume
Hydrophobic interaction chromatography	Differences in surface hydrophobicity
Chromatofocussing	Differences in isoelectric point
Affinity chromatography	Based on biospecific interactions between protein and resin
Hydroxyapatite chromatography	Based on interactions between protein and a calcium phosphate-based media
Reverse phase chromatography	Differences in solubility (surface hydrophobicity)

Adapted from Walsh [2007b] and Forciniti [2008]

Ion exchange chromatography (IEC) is one of the most frequently employed protein purification techniques in the biotechnology industry [Davies and Demain 1999; Flickinger and Drew 1999]. It is based on the reversible electrostatic interactions between the charged amino acid residues on the protein's surface and the charged groups chemically linked to the resin. The chromatographic media has either positively charged groups (anion exchanger) or negatively charged groups (cation exchanger) attached. The type of ion exchanger employed will depend on the particular protein and may be determined through bioinformatic-based predictive approaches or empirically [Flickinger and Drew 1999; Roos 1999]. This method

requires minimal salt concentration during starting conditions, i.e. protein binding, which often necessitates buffer exchange prior to IEC [Roos 1999]. Elution of the absorbed proteins demands a second buffer with potentially high salt concentrations or a change in pH [Davies and Demain 1999; GE Healthcare 2013]. Therefore, column regeneration between processes is needed and can be time-consuming in large-scale production [Roos 1999]. However, IEC achieves high level resolutions, features high capacity due to porous resin increasing the surface area and facilitates straightforward scale-up. Additionally, the process is relatively inexpensive and easy to use. This ensures IEC's continued use for industrial DSP. [Walsh 2007b].

Hydrophobic interaction chromatography (HIC) depends on the interactions between the hydrophobic areas on the protein's surface and the (moderately) hydrophobic groups attached to the chromatographic media. The hydrophobic character of a protein is enhanced with high salt concentrations in the binding buffer and a second buffer is employed during elution [Forciniti 2008]. The exact mechanism of interaction is complex and not yet fully understood which affects optimisation requirements. However, HIC achieves high resolution, yielding good purification factors at optimal conditions [Flickinger and Drew 1999]. It is easy to perform with reproducible results and can be automated allowing fast separation protocols. Drawbacks for this technique include the high manufacturing costs of HIC media and regeneration of the resin may be difficult. However, the chromatographic resin is stable and withstands many cycles making this an economically feasible method during large-scale DSP [Jacob 1999].

In affinity chromatography (AF) the resin's surface is covered with compounds (ligands) that bind a particular protein or family of proteins [Forciniti 2008]. This technique shows very high selectivity due to these specific interactions resulting in a high purification power for AF. A variety of ligands may be utilised such as substrates/inhibitors for enzymes, antibodies for antigens or vice versa, receptors for hormones and lectin for glycoproteins [Walsh 2007b; Forciniti 2008]. Pseudobiospecific ligands such as metal ions or dyes may also be used as they are recognised by specific amino acids [Ratledge and Kristiansen 2001]. The method involves the binding of the target protein with the ligand while all other molecules are removed during a wash step. A second buffer with an altered chemical

composition is employed to elute the bound proteins either due to a change in pH, ionic strength or presence of a competing ligand. There are number of disadvantages to AF including poor stability and high manufacturing costs of biospecific ligands. Additionally, the ligand binding may be chemically complex, hazardous, time consuming or even be non-specific [Subramanian 2012]. However, AF's high selectivity as well as high yields often allows a reduction in the number of steps necessary for the purification of a target protein, decreasing overall production costs while increasing overall yield [Hubbuch and Kula 2007; Walsh 2007b]. The addition of purification tags via genetic engineering to enable the use of AF and simplify the separation process is common practise in laboratory environments. Widely used tags include a calmodulin-binding peptide, the cellulose-binding tag, glutathione *S*-transferase (GST), a small peptide of six histidine residues (His₆), the maltose-binding protein (NusA) and the streptavidin friendly biotin [Bonetta 2006]. Amino (N)-terminal His₆-tags are often the first choice and most commonly observed in laboratory settings. Using this N-terminal tag with *E. coli* expression ensures that the bacterial transcription and translation machinery encounter 5' and N-terminal sequences compatible with RNA synthesis and protein expression, respectively. Recombinant proteins with His₆-tags are easily purified with immobilised metal affinity chromatography (IMAC) [Gräslund *et al.* 2008]. This AF technique employs chelating groups that immobilise transition metals such as copper, cobalt, iron, nickel and zinc onto the surface of the chromatography resin which bind to specific amino acids, in particular histidine residues [Forciniti 2008]. The use of IMAC in combination with the His₆-tag enables purification of proteins with unknown properties. Also, metal chelate ligands are more resistant to physical, chemical and microbial degradation than other AF ligands and usually less expensive [Waugh 2005; Kuo and Chase 2011]. An additional advantage of the His₆-tag is its relatively small size which ensures that it usually does not affect the protein's solubility, only rarely affects its characteristics and does not have a consistent impact on the N-terminal structure of the recombinant protein [Gräslund *et al.* 2008]. Its small size also means that it has a low metabolic burden on the expression strain [Waugh 2005]. A drawback with IMAC purification is the potential metal leakage into the protein solution. Furthermore, the specificity of this technique is not as high as other AF methods [Waugh 2005; Forciniti 2008].

Purification of His₆-tagged recombinant proteins via IMAC is an efficient research tool, however, for the production of therapeutic proteins complete tag removal is essential. It may not be possible to exclude interference of the purification tag with the biological function of the recombinant protein but the necessary tag removal for biopharmaceuticals makes this process expensive, complex and problematic for large-scale use [Cramer and Holstein 2011; Kuo and Chase 2011].

1.5 *E. coli* β -galactosidase

For this research, a model *E. coli* strain capable of producing a genetically engineered protein was designed. The strain allowed production of the recombinant enzyme (β -galactosidase) via standard fermentation and purification methods, while environmental impacts of various processes were evaluated. The *E. coli* β -galactosidase enzyme was selected as a model recombinant protein for this study as it has been reviewed in detail and standard simple activity assays are well established [Lederberg 1950; Wallenfels and Weil 1972]. The following sections will give an overview of the enzyme.

The product of the Z gene within the *lac* operon (*lacZ*) of the *Escherichia coli* genome is the hydrolysing enzyme β -galactosidase (EC 3.2.1.23) which has been extensively studied and characterised in molecular biology and biochemistry and discussed in accompanying articles [Lederberg 1950; Hu *et al.* 1959; Craven *et al.* 1965; Fowler and Zabin 1977; Fowler and Zabin 1983; Jacobson *et al.* 1994; Roth and Huber 1996b; Nichtl *et al.* 1998; Juers *et al.* 2001; Matthews 2005]. This enzyme was central to the development of the operon model by Jacob and Monod [1961] as well as the understanding of gene expression and regulation [Ullmann 2001].

β -Galactosidase participates in the carbohydrate exchange of an organism by catalysing the hydrolytic and transgalactosidic reactions of β -D-galactosides [Roth *et al.* 2003]. The *E. coli* enzyme is heat stable in the presence of β -mercaptoethanol and retains 50 % of its activity after 40 min at 55 °C. No loss of enzymatic activity is observed when β -galactosidase is stored between pH 6 and pH 8 for 30 min at 40 °C. However, the stability decreases greatly below pH 6 and above pH 8 and all biological function is lost at pH 3.5 or lower and pH 11.5 or higher. The treatment of urea, guanidine hydrochloride and mercurials also causes a loss of activity due to

dissociation of the enzymes subunits but may be recovered when the denaturant is removed [Kuby and Lardy 1953; Wallenfels and Weil 1972; Roth and Huber 1996a].

The activity of the enzyme can be readily assayed using synthetic chromogenic, fluourogenic and luminogenic substrates specifically designed for β -galactosidase [Flickinger and Drew 1999] with ortho-nitrophenyl- β -D-galactopyranoside (ONPG) most frequently employed. This substrate is broken down to ortho-nitrophenyl (ONP) and the reaction can be monitored and quantified spectrophotometrically at a wavelength of 420 nm [Lederberg 1950]. The enzyme is often used in molecular cloning as the amino (N)- and carboxyl (C)-terminal regions do not need to be carried on the same DNA molecule to generate β -galactosidase activity (α -complementation). It is possible to monitor insertion of foreign DNA sequences into cloning vectors that have the N-terminal fragment of β -galactosidase as the bacteria will lose this enzymatic activity if there is an insert [Sambrook and Russell 2001].

1.5.1 Structure of *E. coli* β -galactosidase

Fowler and Zabin [1978] determined the amino acid sequence of the protein using enzymatic cleavage with trypsin, chymotrypsin or cyanogen bromide followed by chromatographic separation of the various resulting peptides. The peptides were then hydrolysed and the obtained amino acids identified employing thin layer chromatography [Fowler and Zabin 1978]. The sequence was confirmed with a few minor changes in 1983 by DNA sequencing of the *lacZ* gene [Kalnins *et al.* 1983]. The structure of the *E. coli* β -galactosidase enzyme was determined in 1994 employing X-Ray diffraction crystallography. Using a monoclinic crystal form it was possible to elucidate the structure to 2.5 Å resolution (PDB accession number 1BGL) [Jacobson *et al.* 1994]. Subsequently, this was enhanced to 1.7 Å resolution utilising an orthorhombic crystal form with a smaller unit cell (PDB accession number 1F4H). These structures showed that β -galactosidase is a 464 kDa homotetramer with 222-point symmetry and is roughly ellipsoidal with dimensions of 175x135x90 Å along the twofold axes [Juers *et al.* 2000]. The tetramer has four identical polypeptide chains (labelled A-D in Figure 1.3), each one consisting of 1023 amino acids. Each monomer is made up of five well-defined structural domains (1-5), which are coloured blue, green, yellow, cyan and red in Figure 1.3 and an N-terminal extension consisting of the first approximately 50 residues [Matthews 2005].

Seven apparent cis-peptide bonds were observed and the residues involved all occur within domains at turns or other breaks in the secondary structure [Juers *et al.* 2000]. In the active enzyme the four monomers are arranged around three mutually perpendicular twofold axes of symmetry, considered to be three distinct interfaces relating different pairs of monomers with one another. The ‘active’ and ‘long’ interface can be seen in Figure 1.3 [Matthews 2005].

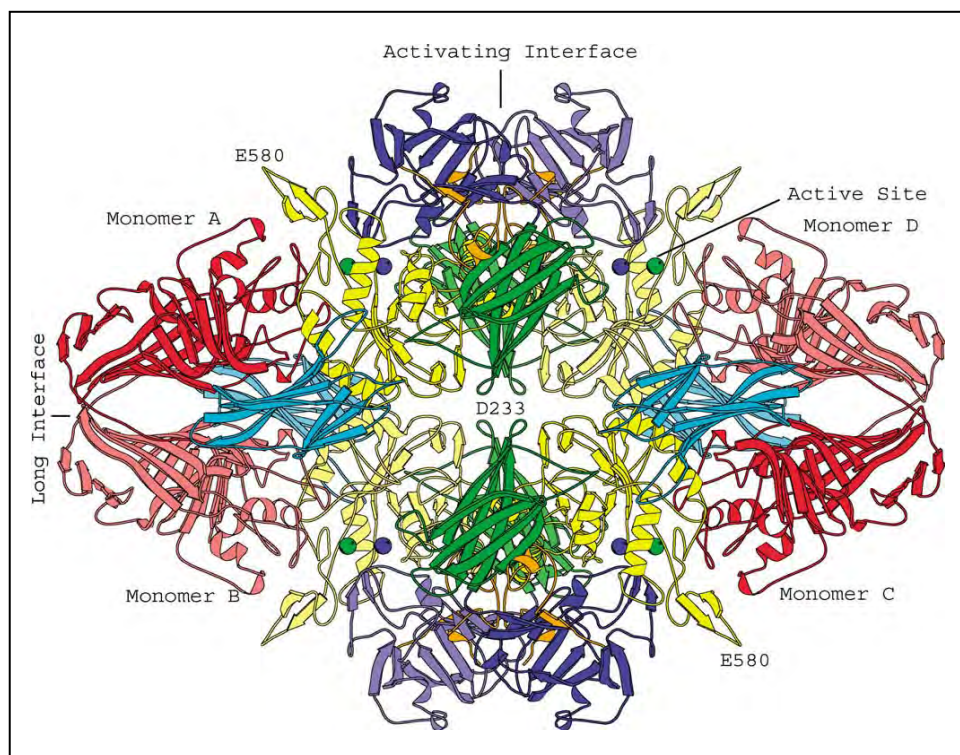


Figure 1.3 View of *E. coli* β -galactosidase tetramer identifying the twofold axes

*Colouring is by domain: N-terminal extension (complementation peptide), orange; Domain 1, blue; Domain 2, green; Domain 3, yellow; Domain 4, cyan; Domain 5, red. Lighter and darker shades of a given colour are used to distinguish the same domain in different subunits. The metal cations in each of the four active sites are shown as spheres: Na^+ , green; Mg^{2+} , blue [Juers *et al.* 2000].*

As the modular structure of each monomer suggests, the folding of this long polypeptide chain may occur due to each domain acting as an independent folding unit [Matthews 2005]. Further investigations determined that the folding of structured, association-competent monomers occurs quickly. The first rate-limiting step is the formation of bi-molecular organisation of those monomers to dimers. The next step, also rate-limiting, is the rearrangement of those dimers, which creates the activation interface necessary for tetramer formation. The last step in the β -

galactosidase folding process is the association of the rearranged dimers to the active tetramer. This step as well as the first step in the pathway is very fast [Nichtl *et al.* 1998]. The folding pathway is represented in Figure 1.4.

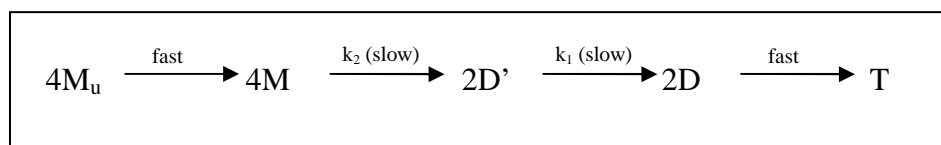


Figure 1.4 Equation representing folding of *E. coli* β -galactosidase

*The four unfolded monomers (M_u) associate quickly to four structured monomers (M). The association of partially folded dimers (D') is a second-order reaction (k_2 rate constant of association) and the conversion of those dimers a first-order reaction (k_1 rate constant of association). T is the active enzyme [Nichtl *et al.* 1998].*

Various studies have shown that specific residues of different segments in the polypeptide chain found close together located around a deep pit formed at the C-terminal end of a triosephosphate isomerase (TIM) barrel are important for the catalytic function of the enzyme [Cupples and Miller 1988; Cupples *et al.* 1990; Ring and Huber 1990; Gebler *et al.* 1992; Juers *et al.* 2001]. This pocket was therefore identified as the substrate binding site [Jacobson *et al.* 1994]. β -Galactosidase' active site intrudes into the central core of a TIM barrel. Portions of loops of the first, second and fifth domain are recruited to stabilise the active site and confer specificity for the disaccharide substrate [Juers *et al.* 2000]. The enzyme has a total of four functional active sites each comprising of segments of two monomers and since it takes two monomers to complete one active site, individual monomers are presumably inactive [Matthews 2005]. There is a continuous channel system along the surface and within the protein which is accessible to bulk solvents. The four active sites can be found at the bottom of such surface channels [Juers *et al.* 2000].

The enzyme is tolerant to deletions and substitutions of amino acids at its N- and C-termini. Up to 26 amino acids can be removed from the N-terminus and replaced with several hundred or more residues of a variety of other proteins without effecting enzymatic activity [Brickman *et al.* 1979; Fowler and Zabin 1983]. This feature is being used to produce β -galactosidase fusion proteins for improving production of recombinant proteins in *E. coli*. This strategy can reduce toxicity, increase yield, improve purification and allow monitoring of expression of the fused protein [Corchero *et al.* 1996].

1.5.2 Mechanism of action

E. coli β -galactosidase requires mono- and divalent cations for its full catalytic activity. Magnesium ions (Mg^{2+}) or manganese ions (Mn^{2+}) activate the enzyme by 5-100 fold and monovalent ions, in particular sodium (Na^+) and potassium (K^+), by 0.3-6 fold, depending on the ion [Wallenfels and Weil 1972; Juers *et al.* 2001]. Using X-Ray data, one Mg^{2+} was identified in each active site and has been shown to have a catalytic and substrate binding role in combination with Glu416, His418, Glu461 and three waters [Juers *et al.* 2001; Wheatley *et al.* 2012]. Structural studies also elucidated five putative Na^+ -binding sites, one of which is in the vicinity of the active site and its removal is likely to perturb the region and reduce activity. During studies determining the mode of action of β -galactosidase it was established that the active site Na^+ , coordinated by Asp201, Phe601, Asn604 and water molecules, is directly involved in the binding of ligands [Juers *et al.* 2000; Juers *et al.* 2001; Wheatley *et al.* 2012].

β -Galactosidase has three activities which will result in the complete breakdown of lactose, a disaccharide. First, lactose is cleaved by the enzyme into galactose and glucose [Juers *et al.* 2001]. The enzyme may also act as a transglycosylase and convert lactose to allolactose, the natural inducer of the *lac* operon. Therefore, *E. coli* can both respond to and utilise lactose as a food source. [Matthews 2005]. The third activity is the hydrolysis of allolactose to galactose and glucose [Juers *et al.* 2001]. The mechanism of action is illustrated in Figure 1.5.

Lactose hydrolysis by β -galactosidase occurs via a two-step mechanism (double displacement) involving a covalent galactosyl-enzyme intermediate [Juers *et al.* 2003]. Crystal structures have identified two conformations, termed “open” and “closed”, for the β -galactosidase enzyme. When there is no substrate present as well as when substrates are bound at the surface of the active site (“shallow” mode) a loop (residues 794-803) at the active site and the side chain of Phe601 are in the “open” conformation. The movement of the substrate deeper into the active site (“deep” mode) initiates the reaction. Glu461 acts as the acid catalyst for the first nucleophilic displacement resulting in an oxocarbenium ion-like transition state. The loop conformation changes to the “closed” state. This reaction forms a quasi-stable covalent bond between the anomeric carbon of galactose and Glu537 while glucose

is displaced. This step is called “galactosylation” because the enzyme has become galactosylated. The second displacement, which is similar to the first, uses either water or an acceptor with an alcohol group as the nucleophile and a second oxocarbenium ion-like transition state is formed. During this step, Glu461 functions as a general base catalyst. If water is the acceptor this reaction is called “degalactosylation” because galactose is cleaved from the enzyme. If another acceptor is used (e.g. sugar or alcohol) the product is a galactosyl adduct and the reaction is called “transgalactylation”. The loop reverts back to the “open” state once galactose or an adduct is formed. Allolactose is formed via this transgalactylation process [Dugdale *et al.* 2010; Wheatley *et al.* 2012].

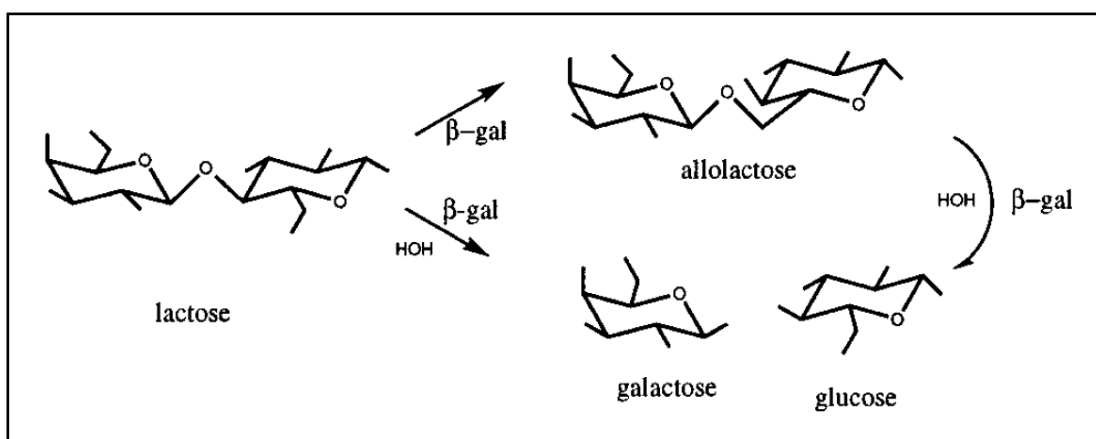


Figure 1.5 General scheme of *E. coli* β -galactosidase action on the natural substrate lactose

The enzyme (β -gal) can either perform hydrolysis (lower path) or transglycosylation (upper path) [Juers *et al.* 2001].

Jurado *et al.* [2002] showed that the kinetic mechanism responds to a model of inhibition by the product (galactose). It was also verified that the enzyme has similar affinity to both substrate and product as both compounds are capable of occupying the active site of the enzyme with equal probability [Jurado *et al.* 2002].

The enzyme shows high specificity to the galactosyl moiety of substrates during binding but is promiscuous to the remainder of the molecule. This particular characteristic has allowed the development of a number of substrate analogues that incorporate a chromophore facilitating the determination of enzyme activity via distinct changes in colour [Juers *et al.* 2001].

1.6 Environmental implications

Assessing the environmental impact of biopharmaceutical manufacturing has become an integral part of process design [Dunn *et al.* 2010]. The production of recombinant proteins as well as cleaning, decontamination and sanitisation processes undertaken in between production runs generate a number of waste streams consisting of biological or chemical waste. The biological wastes are shown in Figure 1.2 (section 1.4.2) and comprise of cellular debris, nucleic acids, host cell proteins as well as truncated, clipped or chemically modified forms of the target protein [Hubbuck and Kula 2007]. The main concern regarding this type of waste is the possible accidental release of recombinant DNA and environmental protection is achieved by waste sterilisation techniques (decontamination and inactivation) [NIH 2002; Walsh 2007a].

The chemical wastes produced include the buffer components employed throughout the purification and concentration of the product protein. DSP buffers consist of mainly water, inorganic salts, bases and acids and the exact combination and material quantities are very specific to each process. Detergents, specific alcohols or organic solvents may also be used during some of the DSP stages. Chemical wastes can significantly contribute to the pollution load of the wastewater of a biopharmaceutical company [Dunn *et al.* 2010]. These waste streams may also include unutilised nutrients from the spent fermentation media which often contains rich components including nitrogen (N) and phosphorus (P) sources [Cliffe *et al.* 2010]. N and P compounds can have serious effects on the environment in particular in terms of nutrient enrichment and resource depletion [Khan and Ansari 2005; Cordell *et al.* 2009]. The environmental effects of phosphorus are discussed in detail in section 1.6.1.

1.6.1 Phosphorus in the environment

Phosphorus (P) is an essential element for all living organisms as it is connected to cell physiology and biochemistry [Ott and Rechberger 2012]. It is a component of nucleic acids, proteins, phospholipids and adenosine-5'-triphosphate (ATP), the primary carrier of chemical energy in biological systems. Therefore, it plays an important role in energy distribution, helps main the pH, is involved in storage and

transfer of coded genetic information and activates many catalytic proteins. Also, inorganic phosphates are a key structural material in vertebrate bone [Novosad 1994; Khan and Ansari 2005]. P is essential for living organisms and in humans too little or too much can cause health problems such as kidney damage, osteoporosis and hypophosphatemia [Knochel 2006].

The earth's crustal rock contains approximately 0.12 % (w/w) of P, making it the 11th most abundant element. Almost all terrestrial P is found in the form of minerals including apatites (chloro and fluoro), vivianite, wavellite and phosphorites. The largest reserve of P occurs in the form of apatite minerals ($\text{Ca}_5(\text{PO}_4)_3[\text{F}, \text{OH} \text{ or } \text{Cl}]$) which are generally insoluble in water [Hanrahan *et al.* 2005; Tiessen 2008]. Readily available P rock deposits are being depleted due to the extensive use of P in agriculture and in products such as detergents, matches, grenades and flares [Massey *et al.* 2010].

In nature, P occurs almost solely in its fully oxidised forms as inorganic phosphates or organic phosphate esters [Tiessen 2008]. These P-compounds circulate through the hydro-, bio- and lithosphere, however, there is no stable gaseous forms of P, making the P matter cycle endogenic [Manahan 2005]. The primary inorganic P cycle consists of three stages through which P circulates very slowly. During the first step phosphate leaches from rocks and soils via weathering and advances to the sea through rivers. The phosphate precipitates in sea as calcium phosphates which are deposited mostly on continental shelves. The last step is the geological uplifting of these oceanic sediments and exposure to weathering [Novosad 1994]. Secondary cycles, represented in Figure 1.6, integrate the movement of P through the life cycles which consists of the uptake of inorganic P by plants, algae and microorganisms followed by the conversion to organic P. These P-containing compounds are transported to animals via nutrient consumption. Decomposition of organic matter will return P to the environment. Microbes link living and non-living P reservoirs by aiding weathering, mineralisation and solubilisation of non-bioavailable P sources allowing conversion into soluble biologically available P [Manahan 2005; Mackey and Paytan 2009]. P is delivered to the aquatic biosphere via natural processes such as weathering of rocks, leaching of soil and through air in form of particulates. However, agriculture, industry and domestic sewage have greatly modified the

global P cycle as they contribute extensively to the P discharge into aquatic systems [Valsami-Jones 2004].

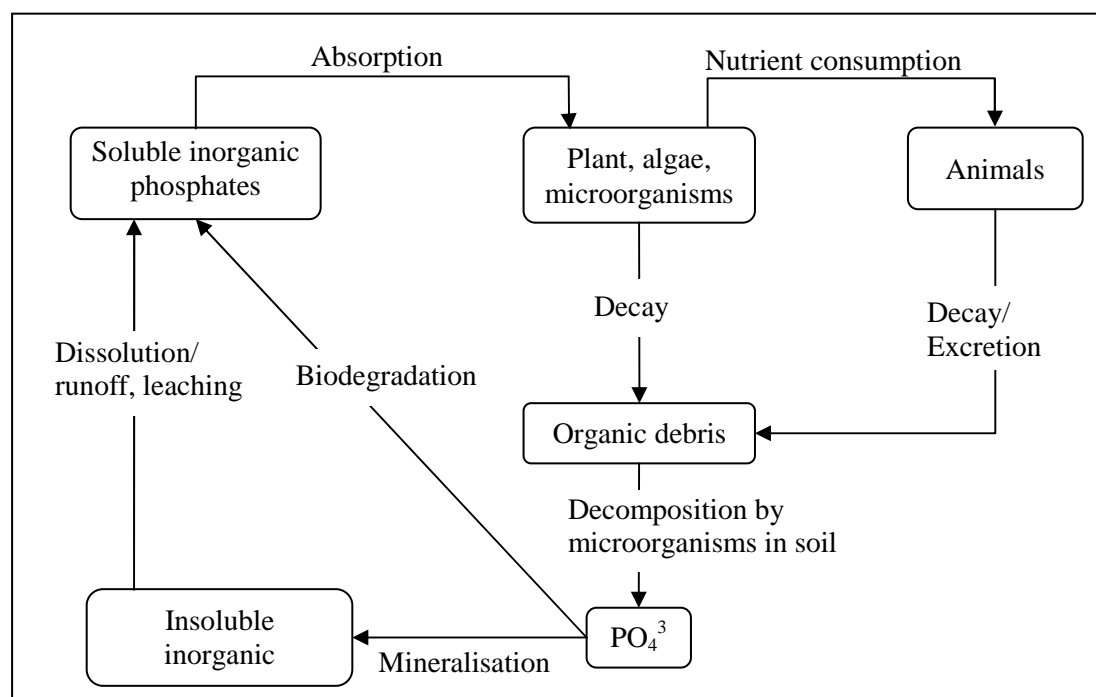


Figure 1.6 The phosphorus cycle

This figure was adapted from Novosad [1994] and Manahan [2005] and shows a simplified phosphorus cycle. Soluble inorganic P may be taken up by plants, algae or bacteria and assimilated into biological phosphorus such as nucleic acids, ATP and phospholipids. Plants, algae and bacteria are consumed by animals thereby absorbing the organic P from photosynthetic organisms. Inorganic phosphorus is returned through decomposition of animal excrements as well as of dead plants and animals by microorganisms. Mineralisation of phosphate returns it to the rock and soil from where it may dissolve through weathering or leaching, to be taken up again by plants, algae or microorganisms.

Naturally occurring P forms include: (a) mineral forms, (b) organic forms such as nucleic acids, phosphoproteins, phosphopolysaccharides, nucleotide cofactors and phosphonates, (c) dissolved inorganic forms (e.g. pentavalent, trivalent or univalent), (d) gaseous forms in the $-III$ oxidation state (phosphine) and (e) particulate or colloidal forms [Hanrahan *et al.* 2005]. The majority of known P compounds contain linkages to oxygen, carbon, nitrogen and/or metals and a significant minority will be bonded to boron, silicon, sulphur or halogens. The biologically important P forms are inorganic P, with P–O or P=O linkages, or organic P, with one, two or three P–O–C linkages forming mono-, di- and triesters, respectively [Valsami-Jones 2004].

1.6.2 Phosphorus in the aquatic environment

The determination of the P species present in water bodies provides valuable information for assessing the health status of the ecosystem, investigation of biogeochemical processes and for supervising compliance with legislation [Worsfold *et al.* 2005]. P occurs in various particulate and dissolved forms in aquatic environments as shown in Figure 1.7. The operationally defined P fractions are classified into non-filtered and filtered species including total P (TP), total reactive P (TRP), total filterable P (TFP) and filterable reactive P (FRP). TP refers to total dissolved and particulate P while TFP and FRP is a measure of the P present in the filtrate of a water sample passed through a membrane filter (typically 0.45 μm) [Maher and Woo 1998; Hammer and Hammer 2004; Worsfold *et al.* 2005].

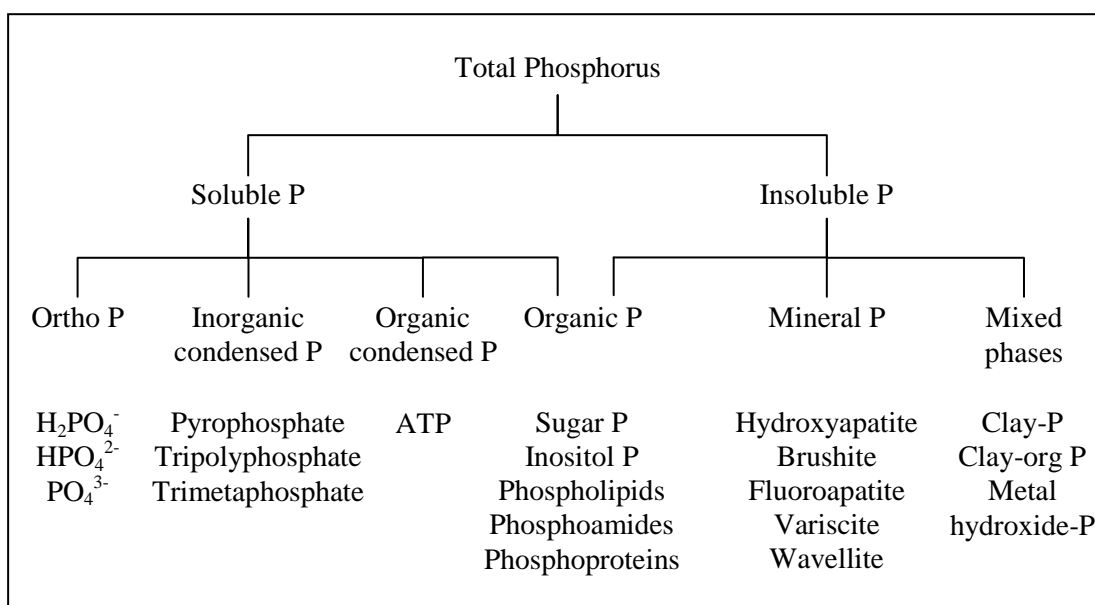


Figure 1.7 Examples of soluble and insoluble phosphorus species found in natural waters

Adapted from Maher and Woe [1998]

Phosphorus exists in natural waters and wastewaters almost exclusively as phosphates and can be classified as condensed phosphates (pyro-, meta- and other polyphosphates), organically bound phosphates, orthophosphates and phosphonates. These compounds may occur in solution, in particles or detritus and in the aquatic biota [Eaton and Franson 1995; Maher and Woo 1998]. The P species most commonly encountered in dissolved forms are outlined in Figure 1.7. Orthophosphates are in an oxidised state (H_2PO_4^- , HPO_4^{2-} , and PO_4^{3-}) and its form is pH dependent. This type of P is readily available for biological utilisation and is

therefore regularly applied to agricultural and residential cultivated land as fertilisers [Maher and Woo 1998]. Storm runoff and to a lesser extent the melting of snow carries these orthophosphates to surface waters [Eaton and Franson 1995]. Organic P compounds entering aquatic systems are usually esters of orthophosphoric acid, i.e. contain C–O–P bonds, and phosphonates (C–P). Those compounds are primarily generated via biological processes, i.e. breakdown of organic matter or excretion of metabolic products, or enter via sewage effluent containing body wastes and food residues [Maher and Woo 1998]. These waste streams contain organic phosphate materials such as nucleic acids, phospholipids and sugar phosphorus compounds [Hammer and Hammer 2004]. Organic condensed phosphates such as adenosine-5'-triphosphate (ATP) are produced by all living organisms and as a result are found in all aquatic ecosystems. Condensed inorganic phosphates with P–O–P linkages are produced for the use in detergents and discharged with domestic and industrial waste streams [Maher and Woo 1998]. In aquatic environments condensed phosphates are slowly converted to orthophosphates [Khan and Ansari 2005].

1.6.2.1 Eutrophication

The primary environmental concern of P release into water bodies is the growth of aquatic plants causing eutrophication, in particular when P is found in conjunction with nitrogen (N) sources. Eutrophication is defined as the enrichment of water by nutrients causing an accelerated growth of algae and higher forms of plant life to produce an undesirable effect resulting from anthropogenic enrichment by nutrients [Csathó *et al.* 2007]. It is not caused by P alone but by an intricate interaction between N, P, environmental conditions, such as temperature, salinity and light, and the physical and hydrological conditions of the surface water [Pierzynski *et al.* 2000]. The first step in eutrophication is the addition of nutrients caused by point source pollutants, which refers to identifiable localised sources primarily inadequate municipal solid and wastewater treatment and disposal, and/or diffuse/non-point source pollutants such as surface and sub-surface run-offs from land [Manahan 2005; Csathó *et al.* 2007]. A gradual addition of nutrients to the surface water will lead to a shift in the aquatic system from oligotrophic, through mesotrophic, eutrophic and eventually to hypertrophic [Pierzynski *et al.* 2000]. This is a slow process, naturally oligotrophic waters, for example, may take thousands of years to become eutrophic,

however, this process is hastened by human activities which elevates the rate of nutrient inputs [Khan and Ansari 2005].

Eutrophication leads to notable changes in the water quality negatively affecting ecological, economic and animal and human health perspectives. This includes problems in water quality, taste, odour and impairment for industrial and recreational use [Pierzynski *et al.* 2000; Khan and Ansari 2005].

In natural waters with little or no pollution, P is frequently a growth-limiting nutrient and therefore often inhibits primary productivity [Crouzet *et al.* 1999; Hanrahan *et al.* 2005]. However, P is used extensively in fertilisers and other chemicals which commonly leads to an accumulation to high concentrations in the water bodies around agricultural fields or densely populated areas due to discharge of human waste [Khan and Ansari 2005]. A number of different organism groups contribute to the primary production in aquatic ecosystems including emergent plants, aquatic bryophytes, epiphytic green and blue-green algae, benthic green and blue-green algae and phytoplankton. These producers require P and N to convert carbon dioxide into organic carbon and diminution of one or both of these nutrients can limit algal growth. Consequently, increased concentrations of P and/or N may stimulate the growth of these aquatic organisms [Nijboer and Verdonschot 2004]. P concentrations as low as $10 \mu\text{g l}^{-1}$ can cause eutrophication processes and result in a deterioration of the water quality [Berg *et al.* 2006].

Additional factors such as current velocity, seasonal patterns, light, temperature and pH may also have an impact on primary production as these can directly influence nutrient uptake, production rates, photosynthesis, respiration and diversity of the aquatic flora [Stevenson *et al.* 1996; Francoeur *et al.* 1999; Khan and Ansari 2005].

During their growth phase primary producers serve as oxygen-generating sources and can lead to an increase in fish population and overall biodiversity of the water system [vanLoon and Duffy 2000]. However, excessive concentrations of algae can reduce the available light and oxygen levels causing adverse ecological consequences [Struijs *et al.* 2011]. A fundamental process in aquatic ecosystems is the decomposition of detritus which is a mixture of dead and living matter consisting of organic material, fungi, bacteria, algae, invertebrates and fishes. Decomposition

rates are usually higher in eutrophic waters than in oligotrophic waters [Nijboer and Verdonschot 2004]. An increased level of primary producers as well as degradation of detritus leads to rapid oxygen depletion in the water body. Low oxygen content in combination with variations in the community food web can seriously threaten aquatic organisms and may result in fish kills [vanLoon and Duffy 2000; Nijboer and Verdonschot 2004].

Additionally, eutrophication can also affect animal and human health as well as the recreational use of the water bodies. Cyanobacteria (blue-green algae) or their toxins may cause health concerns such as eye irritation, skin rash, fever, vomiting, diarrhoea, pneumonia, gastroenteritis, hepatoenteritis and muscular cramps [Hitzfeld *et al.* 2000]. These health problems have been reported in a number of countries as a result of blooms of toxic cyanobacteria in water storage reservoirs, lakes and rivers, following consumption of contaminated drinking water or after recreational contact [Chorus and Bartam 1999].

While a water body undergoing eutrophication can be treated, the necessary control measures may be extensive and difficult to administer. Also, excessive algae and aquatic plant growth can cause taste and odour problems that persist even though the water is treated and filtered prior to use. The water treatment can become expensive and time-consuming for eutrophic water bodies [Ryding and Rast 1989].

The Irish Environmental Protection Agency (EPA) has identified eutrophication as the main threat to water quality in Ireland [EPA 2010]. Although nitrogen and carbon are also essential to the growth of aquatic biota there is great difficulty in controlling the exchange of nitrogen and carbon between the atmosphere and water, and fixation of atmospheric nitrogen by some blue-green algae. Therefore, control of P inputs is critical to reducing freshwater eutrophication [Sharpley 1999]. Phosphorus Regulations require that water quality be maintained or improved by reference to a baseline biological quality rating (rivers) or trophic status (lakes) assigned by the EPA [EPA 2010].

1.6.3 Phosphorus reserves

Phosphorus is obtained from mined rock phosphate (RP) and the majority is integrated with sulphuric acid, nitrogen and potassium to produce mineral fertilisers. In natural systems, P mainly cycles through the plant-soil dynamic, however, agricultural soils require the addition of fertilisers to replenish nutrients lost during harvest [Cordell *et al.* 2009]. Continuous application of approximately 18-52 kg(P) ha⁻¹ per year is necessary to supply agricultural systems [Ott and Rechberger 2012]. Other applications for P compounds include animal foodstuffs, detergents, flame retardants, food additives, glass technology, matches, medicines, metal treatment, nerve gases, oil additives, pesticides, plastics, refractories, smoke generators, surfactants and water treatment [Novosad 1994]. Reserves of RP are found in many countries, but the largest economically recoverable reserves are located in China, United States and Morocco/Western Sahara only [Rosemarin *et al.* 2009]. Therefore, reserves are in control by only those few countries which makes them subject to international political influence. At this time, China has high-quality reserves but imposed a 135 % export tariff, thereby, preventing export to secure domestic supplies [Cordell *et al.* 2009; Vaccari 2009]. Assuming today's rate of extraction the U.S. will run out of RP reserves within the next 30-40 years and already requires RP imports for their fertiliser production [Vaccari 2009]. At current extraction rates, the global reserves are going to run out within the next 50-120 years. However, with an increase in population, improving diets and rising global demand of food and biofuels the extraction rate will increase every year. The demand for P is predicted to rise to 50-100 % by 2050. Also, with quality of RP declining and access more difficult than before, the cost of extraction, processing and shipping are increasing and it may become economically unfavourable to extract P from RP [Cordell *et al.* 2009; Massey *et al.* 2010; Ott and Rechberger 2012]. Further exploration and intensive exploitation of exciting RP sources, including off-shore and/or lower grade deposits are conventional supply-side approaches to the increasing P demands. Other methods comprise of P recovery techniques such as ploughing crop residues back into the soil, composting food waste from households, food processing plants and food retailers and the use of animal and human excreta [Cordell *et al.* 2009]. Additionally, there is growing interest in technical and economical procedures for the recovery and reuse of P from municipal, agricultural

and industrial wastes [Yuan *et al.* 2012]. There are a number of P recovery technologies and processes available such as struvite crystallisation, chemical P removal through the addition of salts (e.g. FeCl_3 , $\text{Al}_2(\text{SO}_4)_3$ or $\text{Ca}(\text{OH})_2$) and biological P removal which relies on a specific group of polyphosphate accumulating microorganisms capable of taking up excessive concentration of P as intracellular storage [Holba *et al.* 2012]. However, as the cost of recovered P is still higher than RP these methods are still far from mainstream practice [Cordell *et al.* 2011].

1.6.4 Integrated Pollution Prevention Control (IPPC)

Any industrial or agricultural activity that has a (high) pollution potential requires an environmental permit. This permit is covered by the European directive 2008/1/EC concerning integrated pollution prevention control (IPPC). The licence is granted by the appropriate authority for each EU member state if certain environmental conditions are met [EUROPA 2011]. In Ireland, the EPA is responsible for granting the IPPC licence and ensures compliance with the directive, its basic obligations and specific requirements. The aims of IPPC are the prevention or reduction of emissions to air, water and land, the reduction of waste as well as warrant the efficient use of energy and resources. The IPPC licence is a single integrated licence which covers all emissions and the environmental management within a company [EPA 2013].

The environmental requirements in permits were supported by applying best available techniques (BAT), introducing emission limit values (ELV), measures for energy efficiency and pollution prevention [OECD 2010]. The EPA and IPPC specify threshold limits for 91 pollutants one of which is total phosphorus (TP). Those limits are registered with the European Pollutant Release and Transfer Register (E-PRTR) which was established through Regulation (EC) No 166/2006 and contains annual data on the main pollutant releases (including P) to air, water and land of about 28,000 industrial facilities across the European Union and EFTA countries. The national TP threshold for release is 5,000 kg per year to water and to land [EEA 2013].

Cliffe *et al.* [2010] showed that a theoretical biotechnology process such as plasmid DNA production may employ up to 1.7 tonnes of P per year during fermentation alone using a modest 10,000 litres batch volume and 52 batches per year. It was also

demonstrated that the majority of exogenous P remains unutilised by the microorganisms. Therefore, the waste streams from plasmid manufacturing can contain more than 90 % of the added P [Cliffe *et al.* 2010].

1.7 Life cycle assessment (LCA)

Traditionally, the focus for industrial process design and optimisation was on economic profit and any environmental impact concerns were approached via pollution management systems. These end-of-pipe technologies dealt with pollutants after they have been formed through filtering or cleaning solutions to reduce the amount of harmful emissions from manufacturing facilities [Li *et al.* 2009]. However, over the last few years more attention has been given to industrial pollution as well as reducing and controlling the environmental impact of factories. As a result of increasing environmental control costs, public concern and more stringent environmental regulations companies are developing procedures such as environmental management systems (EMS) to assist in cleaner production and minimising damage to the environment [Wenzel *et al.* 2000; Li *et al.* 2009]. An EMS is a continual cycle of planning, implementing, reviewing and improving actions that an organisation establishes in order to meet its environmental responsibility [Sage 2009]. The ISO 14000 series is a set of standards which provides a framework to EMS and specifies the requirements in the delivery and operation of an EMS. The subjects covered under these standards can be separated into two areas: organisation evaluation, which covers environmental management system, environmental auditing and environmental performance evaluation, and product evaluation, which includes environmental aspects in product standards, environmental labelling and life cycle assessment (LCA) [Li and Hui 2001]. LCA is a powerful, systematic tool used to estimate the environmental impacts associated with the entire life time of products, processes and services. This incorporates all stages from raw material extraction through material processing, manufacture, distribution, use, repair, maintenance and disposal or recycling [Horne *et al.* 2009]. It is particularly beneficial in determining the relative environmental qualities of alternative products and services. Due to its holistic approach to system analysis, LCA is becoming a vital decision making tool in environmental analysis [Azapagic and Clift 1999b]. Therefore, application of this technique and associated waste minimisation strategies by management, design and

manufacturing can lead to cleaner technologies as well as define and reduce the environmental burden of products [Ciambrone 1997; Manahan 2005].

The LCA process consists of four components: goal and scope definition, inventory analysis, impact assessment and interpretation as illustrated in Figure 1.8. The first step in any LCA is the definition of the system (product, process or activity) under study which is covered by the goal and scope definition phase. During the scope definition the object of the assessment is identified and defined, for example, determination of the effect of P emissions on the environment from large-scale recombinant protein production using *E. coli*.

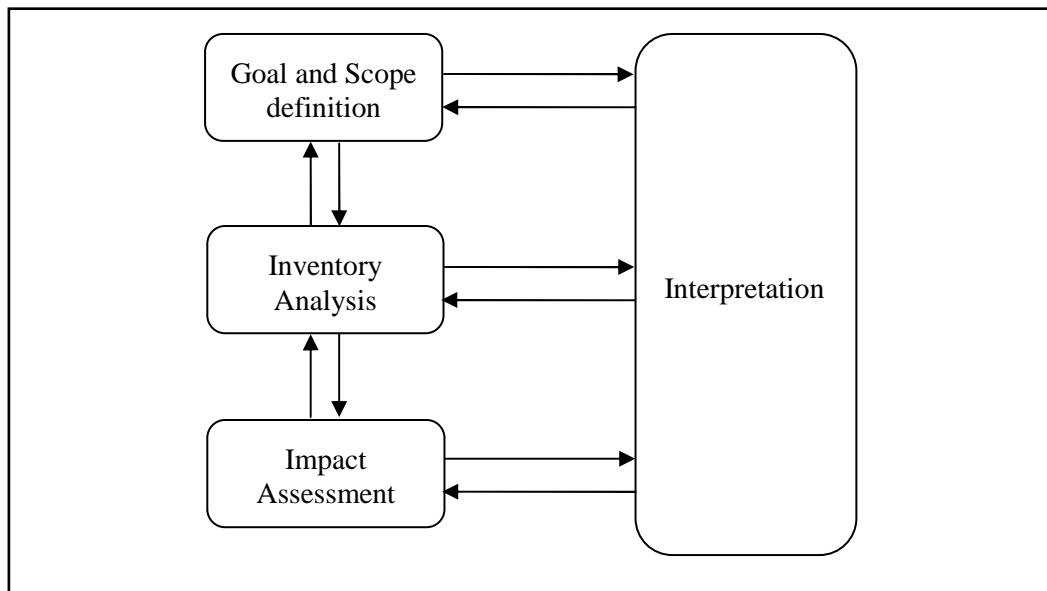


Figure 1.8 Life cycle assessment (LCA) framework

Adapted from SAIC [2006]

This phase also aims to limit the analysis to include that which is significant for the goal of the LCA. It involves establishing the functions of the assessed system, system boundaries, functional unit, data requirements, alternative products or services, key assumptions and limitations of the product [Azapagic 1999; Wenzel *et al.* 2000; Bhandar *et al.* 2003]. The inventory analysis step provides qualitative and quantitative environmentally relevant data regarding the inputs and outputs from the unit and subsequent generation of mass (material) balances [Bhandar *et al.* 2003; Manahan 2005]. The impact assessment phase covers a conversion of the collected data into recognised environmental impact categories followed by an evaluation of

the potential human health and/or environmental risks [SAIC 2006]. This is further divided into three steps:

- Classification (assignment of emission to impact categories)
- Characterisation (calculation of category indicator results for all impact categories)
- Valuation/weighing (to determine which impacts are the most significant)

The interpretation phase analyses the outcome of the previous stages in relation to the goal of the study and identifies key assumptions and data [Bhander *et al.* 2003]. While conventionally LCA was employed to improve the environmental performance of a product, it has been demonstrated that LCA is also a useful tool for process selection, process design and optimisation. It allows balancing the environmental and economic performance, therefore, enabling the choice of best practicable environmental option (BPEO) and best available technology (BAT) [Azapagic and Clift 1999a; Cikankowitz and Laforest 2013].

1.7.1 Environmental evaluation models

In recent years, utilising LCA as a scientific method for evaluating the environmental impacts of products and processes has gained extensive recognition and is widely used in industry in the process of modification, selection and optimisation of product design [Qian *et al.* 2007]. Products are usually manufactured through a series of operational stages and any change in that process will result in a corresponding change in the environmental impact. The theory of unit operation is applied to efficiently assess the environmental impact of a manufacturing procedure. A unit process is specified as a unit action which converts the energy or the physical/chemical/biological state of the inputs into another form [Choi *et al.* 1997]. It consists of inputs, which are raw materials, process and outputs, i.e. product and waste materials. To ensure a precise account of all inputs and outputs of a process, the mass (material) balance principle is applied. It states that the total mass of substances entered into a reaction system is equal to the total mass of substances leaving, i.e. the law of conservation of matter. The mass balance principle can be expressed mathematically as follows:

Total Mass of inputs, $\sum \text{Mass}_{\text{Inputs}}$ = Total Mass of outputs, $\sum \text{Mass}_{\text{Outputs}}$

$$\sum \text{Mass}_{\text{Inputs}} = \sum M_1$$

Where $\sum M_1$ is the total mass of raw materials.

$$\sum \text{Mass}_{\text{Outputs}} = \sum M_2 + \sum M_3$$

Where $\sum M_2$ and $\sum M_3$ are the total mass of the finished product and waste materials, respectively.

Therefore, $\sum M_1 = \sum M_2 + \sum M_3$ [Li and Hui 2001].

The process of quantifying the mass flows of an emission and determining the resulting concentration in the environment is known as assessment modelling. Models are employed as instruments in LCA to describe the relationship between emissions and concentrations and to predict the results of management measures [Van De Meent and De Bruijn 2007]. Models allow a rational interpretation of observations. They not only help in identifying emission relationships in the environment that are not readily obvious but may also provide a means of testing the adequacy of current state-of-the-art theory to describe reality. The results obtained using models can serve as initial guidelines for judging the feasibility of utilising a certain material as well as comparing the alternatives with respect to their environmental and economical qualities [Qian *et al.* 2007]. Modelling provides a way of eradicating the vagueness linked to decision making. Therefore, models are often employed to support decision making in “what if” scenarios to identify the best product design, choice of process or material or the most environmentally friendly technique to use in a manufacturing process [Wenzel *et al.* 2000; Van De Meent and De Bruijn 2007].

1.8 Project aims and overview

Recombinant protein production, in particular biopharmaceutical manufacturing, has become an integral part in the global biotechnology industry. This project aims to analyse the environmental impacts of selected unit processes present in the

manufacturing of these commercial recombinant proteins and reduce the negative environmental impacts without adversely affecting product quantity or quality.

The initial part of this study entails the production of a bacterial strain capable of expressing a functional recombinant model protein. The microorganism selected is *Escherichia coli* MC1061, a *lacZ* deletion mutant, i.e. deficient of endogenous β -galactosidase. Employing molecular biology techniques and optimisation of fermentation conditions will allow the expression of soluble, functional heterologous β -galactosidase with an N-terminal polyhistidine tag (His₆-tag).

Laboratory-scale simulation of fermentation and protein expression using three types of media commonly employed for (engineered) *E. coli* will be conducted. Environmental analysis, specifically determination of total phosphorus, is going to be carried out on the various types of media. Where possible, phosphorus minimisation strategies will be undertaken to aid in the development of an environmentally improved production process. The resulting biomass as well as recombinant product quantity and quality will be established to ensure that there are no adverse effects compared to the original media formulation.

Downstream processing strategies at laboratory-scale for the model protein will be developed using more traditional multi-chromatographic purification techniques, i.e. a combination of ion exchange, size exclusion and hydrophobic interaction chromatography, as well as an affinity-based strategy employing immobilised metal affinity chromatography (IMAC). The two purification schemes are going to be compared from an environmental viewpoint in terms of total phosphorus, chemical oxygen demand, ammonia, organic nitrogen and water usage. Additionally, alternative buffers will be tested for purification using the affinity-based scheme and their environmental impact with respect to total phosphorus and chemical oxygen demand will be assessed.

A quantitative mass balance of phosphorus throughout the production processes will be established. The mass balance simulation will allow the investigation of employing phosphorus-reduced media as well as comparing the various purification schemes and detection of pollution “hot spots”.

To further evaluate the environmental impacts of the wastes generated from the process, an environmental model, entitled ecological hazard score (EHS), is going to be utilised. This model produces a single score of each of the waste streams as well as a total score for the process. This allows the identification of the most environmentally friendly waste streams and ultimately unit operations but also entire production process. Comparison of different media, standard and phosphorus-reduced, as well as different purification schemes, will facilitate an evaluation of the recombinant protein manufacturing. These comparisons can permit possible environmental improvement of the process.

Chapter Two – Generation of *E. coli* model strain

2.1 Introduction

An *Escherichia coli* expression system was selected as this bacteria is still the preferred choice for recombinant protein expression at laboratory- and industrial-scale [Chen 2012]. Low cost and simplicity in cultivation together with available design options, standard molecular cloning techniques and production equipment simplifies engineering of the host strain as well as large-scale manufacturing of the recombinant protein [Kamionka 2011]. Quantitatively, around 60 % of all biopharmaceuticals are produced using *E. coli* expression systems [Ransohoff 2010].

The recombinant model protein selected for this study is the *E. coli* derived β -galactosidase enzyme, most commonly known as lactase (E.C. 3.2.1.23). Its molecular biology and biochemistry, including mechanism of action and structure, have been extensively studied and reviewed in detail [Lederberg 1950; Hu *et al.* 1959; Craven *et al.* 1965; Fowler and Zabin 1977; Fowler and Zabin 1983; Jacobson *et al.* 1994; Juers *et al.* 2001; Matthews 2005; Nichols *et al.* 2007; Jancewicz *et al.* 2012; Wheatley *et al.* 2012]. Moreover, this β -galactosidase is a complex protein, its functionally active form being a homotetramer with an overall molecular mass of 465 kDa [Juers *et al.* 2000]. This makes the enzyme a good model for biopharmaceutical proteins which are large (range from 10 kDa upwards), complex molecules [Ho 2010].

The model strain (*E. coli* MC1061) was chosen as it is devoid of a β -galactosidase gene, rendering straightforward the development of an engineered (recombinant) strain capable of producing this enzyme via the introduction of a β -galactosidase coding sequence (*lacZ* gene) derived from the β -galactosidase positive *E. coli* strain W3110.

2.2 Materials

2.2.1 Chemicals and reagents

Biochemical reagents were purchased from Sigma-Aldrich Chemical Co., Dublin, unless otherwise stated. All other chemicals and reagents used were of analytical grade and obtained from standard sources: Sigma-Aldrich Chemical Co., Dublin; Fluka – a subsidiary of Sigma-Aldrich; British Drug House Laboratory Supplies, Pooles, Dorset, U.K. (BDH); May and Baker Laboratory Chemicals Ltd, Dagenham, UK (MandB); Merck, Darmstadt, Germany (Merck) or Lennox Laboratory Supplies Ltd., Dublin (Lennox).

2.2.2 Molecular biology reagents

The molecular biology reagents used during this study are listed in Table 2.1. Kits were used for genomic DNA extraction from *E. coli*, plasmid DNA extraction and to purify the polymerise chain reaction (PCR) product. The kits contained all reagents and method instructions necessary for each technique and no additional reagents were required.

Table 2.1 Molecular biology reagents used during this study

Reagent	Use	Source
DNeasy Blood and Tissue Kit	Genomic DNA extraction	Qiagen Ltd., Crawley, U.K.
QIAprep Miniprep	Plasmid DNA extraction	Qiagen Ltd., Crawley, U.K.
QIAquick PCR Purification Kit	Clean-up of PCR and Restriction Digest product	Qiagen Ltd., Crawley, U.K.
TaKaRa ExTaq DNA Polymerase	Polymerase used during PCR	Lonza biologics plc., Slough, U.K.
dNTP mix	dNTPs used during PCR	Lonza biologics plc., Slough, U.K.
Restriction Enzymes (BamH1 and HindIII)	Restriction digestion of PCR product and purified plasmid	Roche Ireland Ltd., Clarecastle
T4 Ligase	Ligation of PCR product and plasmid	Roche Ireland Ltd., Clarecastle
AcTEV Protease	Cleavage of polyhistidine tag	Life Technologies, Bio-Sciences, Dublin
DNase	Breakdown of DNA	Sigma-Aldrich Chemical Co., Dublin

2.2.3 Molecular strains

Escherichia coli strain K12 substrain W3110 was purchased from the German Collection of Microorganisms and Cell Cultures (DSMZ), Braunschweig, Germany (www.dsmz.de). This strain arose from the original *E. coli* K12 after treatment with UV light followed by selection of a strongly fermenting colony on eosin methylene blue (EMB)-gal media resulting in several point mutations and an inversion of a considerable chromosome segment [Bachmann 1972; Jensen 1993].

E. coli strain DH5⁻ was used for the initial cloning and plasmid propagation and *E. coli* MC1061 was employed during protein expression as it is a *lacZ* deletion mutant. Both of these strains as well as the pProEx-HTb expression plasmid were present in the laboratory collection of Professor Gary Walsh, University of Limerick, Ireland.

The pProEx-HTb expression vector contains the *colE1 ori* origin of replication, the *bla* gene for ampicillin resistance, an N-terminal His₆-tag and a TEV cleavage site. Expression is under the control of the *trc* promoter, which is recognised by the *E. coli* RNA polymerase.

Table 2.2 shows the microbial strains and the plasmid used for this study as well as their genotype and relevant phenotype. Figure 2.1 depicts a map of the pProEx-HTb vector.

Table 2.2 Microbial strains and plasmid used in this study

Strain	Genotype	Relevant phenotype
<i>E. coli</i> W3110	F ⁻ λ ⁻ rph-1 INV(rrnD, rrnE)	Closely related to wild-type strain (<i>E. coli</i> K12)
<i>E. coli</i> DH5	F ⁻ endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80 <i>dlacZ</i> ΔM15 Δ(<i>lacZYA-argF</i>)U169, hsdR17(r _K ⁻ m _K ⁺), λ ⁻	Propagation of plasmids
<i>E. coli</i> MC1061	F ⁻ Δ(<i>ara-leu</i>)7697 [<i>araD139</i>] _{B/r} Δ(<i>codB-lacI</i>)3 galK16 galE15 λ ⁻ e14 ⁻ mcrA0 relA1 rpsL150(<i>strR</i>) spoT1 mcrB1 hsdR2(r ⁻ m ⁺)	Sm ^R , does not express β-galactosidase
Plasmid	Marker	Relevant phenotype
pProEx-HTb	Amp ^R	High-copy number <i>E. coli</i> plasmid

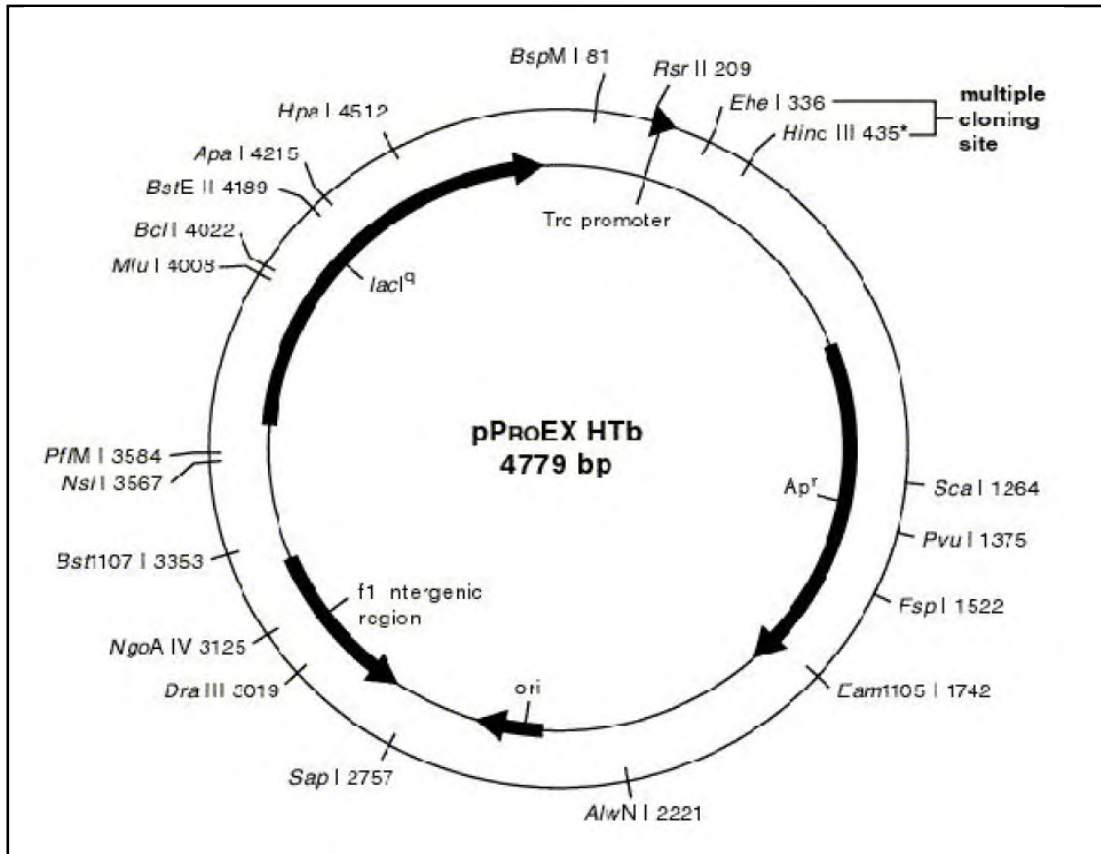


Figure 2.1 Map of pProEx-HTb plasmid [EMBL 2012]

2.2.4 Media and supplements

Culture media and media ingredients were obtained from Sigma-Aldrich Chemical Co., Dublin. Three fermentation media were selected for the study: Luria-Bertani (LB) broth, Terrific broth (TB) and minimal media 9 supplemented with yeast extract (M9/YE). Each media was prepared as per manufacturer's instruction or according to Sambrook and Russell [2001]. Powder media/media ingredients were combined with distilled water and sterilised by autoclaving at 121 °C, 15 psi for 15 min. The compositions of the media used during this study are summarised in Table 2.3. Chemically defined minimal media 9 (M9) was supplemented with 0.2 % (w/v) yeast extract to obtain higher cell densities [Sambrook and Russell 2001].

Table 2.3 Composition of media used during this study

Component (in g l ⁻¹)	Medium		
	LB	TB	M9/YE
Tryptone	10.00	12.00	-
Yeast Extract	5.00	24.00	2.00
NaCl	10.00	-	0.50
Glycerol	-	4 % (v/v)	-
Glucose	-	-	4.00
KH ₂ PO ₄	-	2.31	3.00
K ₂ HPO ₄	-	12.54	-
NaH ₂ PO ₄ ·7H ₂ O	-	-	12.80
NH ₄ Cl	-	-	1.00
MgSO ₄	-	-	0.20
CaCl ₂	-	-	0.01

Composition from Sambrook and Russell [2001]

2.3 Experimental methods

2.3.1 Microbial culturing techniques and cell lysis

Regeneration of the purchased microbial strain and glycerol stocks of strains already present in the laboratory was achieved by inoculation in their designated liquid media (LB broth for all). Details on culture media and growth conditions for the purchased strain were obtained from the website of the culture collection (<https://www.dsmz.de/>). The inoculated liquid media were subsequently incubated at the optimum growth temperature (37 °C) for 16 h. Following incubation, 50 µl of the liquid culture was spread-plated onto LB agar and incubated at 37 °C for 15-24 h. Serial transfer from old to fresh solid media every four weeks achieved maintenance of microbial cultures. Glycerol stocks in LB media containing 15 % (v/v) glycerol for each strain were prepared and stored at -20 °C and -80 °C to maintain viable sources for culturing in the long term [Sambrook and Russell 2001]. Microbial strains with the pProEx-HTb plasmid were grown under selective conditions in the presence of ampicillin (100 µg ml⁻¹).

All media were stored at room temperature and sterilised by autoclaving at 121 °C for 15 min at 15 psi pressure [Prescott *et al.* 2002].

The cell growth was monitored by measuring the optical density at 600 nm (OD₆₀₀) using the UV-visible spectrophotometer (Shimadzu UVmini 1240, Shimadzu Scientific Instruments, Torrance, U.S.).

For cell lysis, cultures of bacteria being harvested were grown until appropriate OD₆₀₀ were achieved (8-16 h) or expression of recombinant protein was completed. These cultures were then centrifuged at 7,500 rpm for 20 min at 4 °C and the pellets were stored at -80 °C for a minimum of 18 h. The cell pellets were defrosted on ice and resuspended in 500 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄, 10 mM MgCl₂, 10 mM β-mercaptoethanol, 50 mg l⁻¹ soybean trypsin inhibitor, pH 7.0. The samples were kept on ice and set up for sonication ensuring that the sonicator probe was immersed in the liquid. All cell solutions were sonicated for 20 min at 20 hrz in 30 s bursts with intermittent cooling periods of 2 min using the Branson SLPe ultrasonic homogenizer (Branson, Danbury, USA). Once sonicated the samples were centrifuged at 10,000 rpm for 20 min at 4 °C to remove cellular debris.

2.3.2 Bioinformatic analysis

DNA and protein sequences required for this study were retrieved from the public databases GenBank and European Molecular Biology Laboratory (EMBL) Bank distributed by the National Centre for Biotechnology Information (NCBI, USA; <http://www.ncbi.nlm.nih.gov/>) and the European Bioinformatics Institute (EBI, UK; <http://www.ebi.ac.uk/>) [Kulikova *et al.* 2007].

The Basic Logic Alignment Search Tool (BLAST) program was used for searching the DNA databases for similar sequences [Altschul *et al.* 1997]. Blast searches were performed on EBI's website: www.ebi.ac.uk/Tools/sss/ncbiblast/nucleotide.html to determine if the designed primers for *lacZ* amplification are unique to the gene of interest.

The physicochemical properties of the designed primers were assessed using the BioMath calculator program designed by Promega which was accessed from

<http://www.promega.com/biomath/calc11.htm> [Nakano *et al.* 1999; Sambrook and Russell 2001]. The BioMath calculator determined the percentage of GC content and the annealing temperatures of the primers.

Nucleotide and protein sequences were aligned using the computer program ClustalW [Larkin *et al.* 2007] which was accessed through the ExPASy website (www.expasy.org). Where necessary, the DNA sequences were translated to their corresponding protein sequence with the help of the ExPASy translate tool.

Compute pI/MW, also accessed through the www.expasy.org website, was employed for theoretical computation of the recombinant protein's isoelectric point (pI) and molecular mass [Gasteiger *et al.* 2005].

2.3.3 DNA isolation and sequencing

Cell cultures of *E. coli* W3110 were grown appropriately and centrifuged at 7,500 rpm, 4 °C to prepare for genomic DNA (gDNA) isolation using the Qiagen DNeasy kit (Table 2.1) according to the manufacture's protocol for the purification of total DNA from gram-negative bacteria. The final purified DNA solution was stored at -20 °C.

High-purity plasmid DNA for the use of cloning and sequencing was isolated from liquid cell cultures using the Qiagen miniprep Plasmid Prep kit (Table 2.1). Final plasmid solutions used for cloning were stored at -20 °C. Plasmid DNA for sequencing (1-2 g) was mixed with 10 volumes of n-butanol and incubated at room temperature for 15 min, followed by centrifugation at 16,000 rpm for 30 min. After removal of the supernatant, the pellet was dried at room temperature for 8-12 h. Sequence analysis was carried out by MWG Biotech (Ebersberg, Germany) using fluorescently labelled standard primers.

During screening of transformants, colonies that appeared on transformation plates after incubation for 10-16 h, were grown in 5 ml liquid cultures overnight at 37 °C. Plasmid DNA was isolated from these strains via the standard alkaline lysis with SDS as described by Sambrook and Russell [2001].

PCR amplicons pre and post restriction enzyme digestion were purified using the QIAquick PCR purification kit, a DNA clean up tool (Table 2.1).

The quality and purity of all extractions was monitored by agarose gel electrophoresis using 0.8 % (w/v) agarose gels stained with ethidium bromide (0.2 g ml⁻¹) according to standard procedures [Sambrook and Russell 2001]. The DNA ladder (New England Biolabs) contained 10 bands with fragments ranging from 0.5 to 10.0 kilobases (kb). Gels were run at 80 V for 1.5-2 h and subsequently scanned with the Gene-genius bioimaging system (Syngene Synoptics Ltd, Cambridge, UK).

The purity of the samples was also assessed by the ratio of absorbancies at 260 nm and 280 nm [Sambrook and Russell 2001].

2.3.4 Polymerase chain reactions

Polymerase chain reactions (PCR) were carried out using 50-100 ng of gDNA from *E. coli* W3110 and the polymerase ExTaq (Table 2.1). The gene and amino acid sequences for β -galactosidase from *E. coli* W3110 have been published on the GenBank database with accession number AP_000996.1 and are shown in Appendix A. The *lacZ* gene is located on the complement strand from nucleotide 362,455 to 365,529 on the genome of W3110 (accession number AC_000091.1) giving a total gene length of 3075 base pairs (bp). Using this data, primers were designed that allowed amplification of the *lacZ* gene (Appendix B) and incorporation of restriction enzyme recognitions sites. The primers are listed in Table 2.4 and the PCR cycle parameters used are tabulated in Table 2.5. Recognitions sites for restriction enzymes (RE) were incorporated into each primer. The forward primer contains the recognition site for BamH1 and the reverse primer for HindIII. This allowed directional insertion of the gene into the pProEx-HTb plasmid.

Table 2.4 PCR primers designed to amplify *lacZ* from *E. coli* W3110

Primer	Sequence	RE Recognition site
lacZ_Forward	GATGGATCCATGACCATGATTACGGATTCCTG	BamH1
lacZ_Reverse	GGGAAGCTTTTATTTTTGACACCAGACCAACTGG	HindIII

The restriction enzyme recognitions sites are underlined in the sequence.

Table 2.5 PCR cycle parameters used for *lacZ* amplification

Temperature	Time	Number of cycles
95 °C	3 min	1
95 °C	30 s	
58 °C	30 s	35
72 °C	1 min	
72 °C	5 min	1

2.3.5 Restriction enzyme digestion of DNA and ligation

The restriction enzymes BamH1 and HindIII (Table 2.1) were selected as there are no recognitions sites for these enzymes within the *lacZ* gene (Appendix C). Both enzymes show optimal catalytic activity at the same buffer conditions allowing a double digest [Roche Diagnostics 2010]. The purified PCR product and isolated pProEx-HTb were restricted as per manufacturer's instruction at 37 °C for 2 h.

Ligation of the restricted gene and plasmid was carried out using T4 DNA ligase (Table 2.1) as per manufacturer's directions. All ligation reactions were carried out at 16 °C for 16 h followed by inactivation of the T4 ligase enzyme by incubating the reaction mixture at 65 °C for 10 min.

2.3.6 Transformation

E. coli strains were rendered competent according to the method of Cohen *et al.* [1972] as described in Sambrook and Russell [2001] with minor modifications.

Briefly, logarithmic growth phase cells were prepared by diluting overnight liquid cultures 1 in 100 and incubating them at 37 °C for 3-4 h until an OD₆₀₀ of 0.5-0.6 was reached. Bacterial cultures were harvested by centrifugation for 10 min at 4,500 x g, 4 °C. The cells were resuspended in 20 % culture volume using 10 mM sodium acetate pH 5.6, 50 mM MnCl₂, 5 mM NaCl and stored on ice for 5 min. Subsequently, the cells were pelleted by centrifugation for 10 min at 4,500 x g, 4 °C. Finally, the bacterial cells were resuspended in 2 % culture volume using 10 mM sodium acetate pH 5.6, 5 mM MnCl₂, 5 % (v/v) glycerol, 70 mM CaCl₂ and 50 µl aliquots of the cell suspensions were stored at -80 °C.

Plasmids were introduced into competent *E. coli* by transformation according to the method of Cohen *et al.* [1972] with minor changes. Competent cells were removed from -80 °C and allowed to thaw on ice. Plasmid DNA/ligation mixture containing 50-100 ng of DNA was added to 30 µl of competent cells in a pre-chilled microfuge tube and gently mixed with the cells. The tube was then stored on ice for 10 min and subsequently placed in a water bath at 42 °C to create a heat shock for 45 s after which the cells were placed on ice for a further 10 min. The transformation mixture was incubated at 37 °C, 100 rpm for 40 min in 500 µl LB broth and afterwards plated on LB agar supplemented with ampicillin. The plates were incubated for 20-24 h and resulting colonies analysed for the presence of the *lacZ* gene.

2.3.7 Optimisation of recombinant protein expression in *E. coli*

E. coli MC1061 bearing an expression plasmid with the *lacZ* insert was grown in LB broth, supplemented with ampicillin, at 37 °C, 250 rpm for 16 h. These cultures were diluted 1 in 100 into fresh LB broth, also supplemented with ampicillin, and incubated at 37 °C, 250 rpm for 3.5-4.5 h to an OD₆₀₀ of approximately 0.6. Isopropyl-β-D-thiogalactopyranoside (IPTG) was added to final concentrations ranging from 0.025-0.5 mM to induce recombinant protein expression. The cultures were incubated for a further 4 h at 37 °C with agitation or for a further 6 or 16 h at 25 °C with agitation. Cells were harvested by centrifugation at 7,500 rpm, 4 °C and stored at -80 °C for a minimum of 18 h.

During optimisation studies the pellet underwent 5-6 freeze/thaw cycles and was then resuspended in 5 % culture volume using 20 mM phosphate, 150 mM NaCl, pH 7.4 followed by incubation on ice for 20 min with 10 µg ml⁻¹ DNase. To separate soluble from non-soluble protein the solutions were centrifuged for 20 min at 10,000 rpm, 4 °C. The supernatant was aliquoted into fresh microfuge tubes and the pellet resuspended in an equal volume with 20 mM phosphate, 150 mM NaCl, pH 7.4. All pellet and supernatant samples were analysed via SDS-PAGE as outlined in section 2.3.8.1.

Once optimal expression parameters were determined the recombinant protein was produced using these conditions in larger culture volumes.

2.3.8 Polyacrylamide gel electrophoresis (PAGE)

2.3.8.1 Gel preparation and running

Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) was carried out according to the method of Laemmli [1970] as described by Sambrook and Russell [2001] using an 8 % gel and a vertical electrophoresis system.

An 8 % resolving gel (total volume 10.0 ml) contained 4.6 ml distilled water, 2.7 ml 30 % acrylamide mixture (29 % (w/v) acrylamide and 1 % (w/v) N, N'-methyl-bis-acrylamide), 2.5 ml 1.5 M Tris-Cl (pH 8.8), 0.1 ml 10 % (w/v) sodium dodecyl sulphate (SDS), 0.1 ml 10 % (w/v) ammonium persulphate (APS) and 0.007 ml N, N, N', N'-tetramethylethylenediamine (TEMED). The 10 % (w/v) APS solution was prepared fresh daily. For the 5 % stacking gel (total volume 3.0 ml) 2.1 ml distilled water, 0.5 ml 30 % acrylamide mix, 0.38 ml 1.0 M Tris-Cl (pH 6.8), 0.03 ml 10 % (w/v) SDS, 0.03 ml 10 % (w/v) APS and 0.003 ml TEMED were combined. The loading buffer contained 50 mM Tris-Cl (pH 6.8), 100 mM dithiothreitol (DTT), 2 % (w/v) SDS, 0.1 % (w/v) bromophenol blue and 10 % (v/v) glycerol. The DTT was added immediately prior to use. The Tris-glycine electrophoresis buffer contained 25 mM Tris, 250 mM glycine and 0.1 % (w/v) SDS.

The molecular mass protein marker (Sigma-Aldrich; SDS6H2) used contained porcine myosin (200,000 Da), β -galactosidase from *E. coli* (116,000 Da), phosphorylase b from rabbit muscle (97,000 Da), bovine albumin (66,000 Da), albumin from chicken egg white (45,000 Da), and carbonic anhydrase from bovine erythrocytes (29,000 Da). In addition, the Mark12 (Life Technologies; LC5677) an unstained protein molecular mass standard was employed during the study which contains myosin (200 kDa), β -galactosidase (116.3 kDa), phosphorylase b (97.4 kDa), bovine serum albumin (66.3 kDa), glutamic dehydrogenase (55.4 kDa), lactate dehydrogenase (36.5 kDa), carbonic anhydrase (31 kDa), trypsin inhibitor (21.5 kDa), lysozyme (14.4 kDa), aprotinin (6 kDa), insulin B chain (3.5 kDa) and insulin A chain (2.5 kDa).

The glass plates were assembled according to manufacturer's instructions and the resolving and stacking gels were allowed to polymerise separately for a minimum of 30 min. Protein samples were combined with loading buffer and heated to 95-100 °C

for 5 min. The protein molecular mass marker was denatured by heating to 95-100 °C for 2 min. The samples and protein marker were added to separate wells and loading buffer was placed into unused wells. Electrophoresis was carried out at 120 V until the bromophenol blue reached the bottom of the resolving gel (2.5-3 h). The glass plate sandwich was subsequently removed from the apparatus, disassembled and the gel extracted.

Protein staining was carried out in accordance with the modified protocol of Sambrook and Russell [2001] using EZBlue staining reagent from Sigma-Aldrich (G1041) as a colloidal stain, which reacts only with the proteins not the gel itself. The gels were washed with distilled water until clear protein bands were visible.

2.3.8.2 Native PAGE and zymogram

Non-denaturing (or native) electrophoresis was carried out as described by Deutscher [1990]. The method is similar to that outlined above (section 2.3.8.1) with the exception that SDS was omitted from the gel and all buffers. In addition, the samples were not heated prior to loading and DTT was not added to the loading buffer. The 5 % resolving gel contained 5.72 ml distilled water, 1.67 ml 30 % acrylamide mix, 2.5 ml 1.5 M Tris, 0.1 ml 10 % (w/v) APS and 0.01 ml TEMED (total volume 10.0 ml) and the 5 % stacking gel is the same as described in section 2.3.8.1 with the exception of SDS which was replaced with distilled water. The native gels underwent electrophoresis at 120 V until the bromophenol blue reached the bottom of the resolving gel (2.5-3 h). The gels were either stained for protein using EZBlue or for β -galactosidase activity.

β -Galactosidase activity staining was performed according to the method described by Chakraborti *et al.* [2000]. After the native gel was electrophoresed it was washed twice with distilled water and then incubated in an x-gal (5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside) solution for 20 min at 50 °C with gentle rotation (50-75 rpm). The x-gal solution was prepared by dissolving it initially in a small volume of dimethyl sulfoxide (DMSO), followed by dilution to 0.02 % (w/v) with 0.2 M sodium phosphate buffer, pH 7.0. After incubation a blue-green band on the gel indicated β -galactosidase activity.

2.3.8.3 Western blotting

Western blotting was carried out to confirm the presence of the His₆-tagged product according to standard methods as described in Sambrook and Russell [2001]. 2.5-20 µg of recombinant protein was loaded per lane on an SDS-PAGE gel to ensure detection with the antibody. Electrophoresis of the gel was carried out as described in section 2.3.8.1. Subsequently, the proteins were transferred onto a PVDF membrane using the Pierce Semi-dry transfer system according to manufacturer's instruction with the Pierce Semi-dry transfer buffer (35035). The membrane was blocked using 5.0 % (w/v) skim milk powder in 20 mM phosphate, 150 mM NaCl, pH 7.4 for 1 h at room temperature or alternatively overnight at 4 °C. The proteins were probed using Monoclonal Anti-polyHistidine Peroxidase conjugate which was afterwards visualised with 3, 3', 5, 5'-Tetramethylbenzidine (TMB).

2.3.9 β-Galactosidase assay

β-Galactosidase activity was assayed based upon the method of Steers *et al.* [1971] with some modifications. The assay system contained 0.4 ml of 5 mM ortho-nitrophenyl-β-D-galactopyranoside (ONPG) in 200 mM sodium phosphate buffer, 10 mM MgCl₂, pH 6.5 and 0.1 ml of enzyme suitably diluted in the same buffer. The reaction was allowed to proceed for 15 min at 50 °C and was stopped by the addition of 0.5 ml of 1.0 M sodium carbonate (Na₂CO₃). Both substrate and enzyme solutions were equilibrated to the assay temperature prior to initiation of reaction. An assay blank contained enzyme and substrate solution, which were incubated separately for the duration of the reaction period and mixed only after the addition of stopping solution to the substrate. Absorbance of the assay solution was measured after cooling to room temperature at 420 nm with a UV-visible spectrophotometer.

A standard curve of ortho-nitrophenol (ONP) concentration (nmol ml⁻¹) versus absorbance at 420 nm was constructed to quantify the amount released during the assay. ONP standard solutions were prepared in triplicate by diluting a stock solution of 10 mM ONP in 200 mM sodium phosphate buffer, 10 mM MgCl₂, pH 6.5. Standard solutions ranged from 50-1250 nmol ml⁻¹. Construction of a standard curve was carried out by mixing 0.5 ml of standard solution and 0.5 ml of stopping solution followed by determining the absorbance values at 420 nm.

One unit of β -galactosidase activity (1.0 EU) was defined as the amount of enzyme capable of releasing 1 nmol of ONP min^{-1} under the defined assay conditions.

2.3.10 Determination of temperature versus activity profiles

Temperature versus activity profiles were established by carrying out the β -galactosidase assay (section 2.3.9) in triplicate at different temperatures for each crude β -galactosidase enzyme. The temperature range used was 25-70 °C. The relative activity at the different temperature values was calculated as a percentage of activity at the optimum temperature. Temperature values versus percentage relative activities were plotted to yield the temperature profile for crude β -galactosidase samples.

2.3.11 Determination of pH versus activity profiles

The crude extract of β -galactosidase producing strains was assayed for activity in triplicate by the standard assay procedure (section 2.3.9) with the exception of sodium phosphate buffer, pH 6.5, being substituted with various buffers in the pH range 3.0 to 10.0. The buffers used were 0.2 M citric acid- Na_2HPO_4 for pH 3.0-7.0, 0.2 M sodium phosphate buffer for pH 6.5-8.0 and 0.2 M boric acid-NaOH for pH 8.0-10.0. The relative β -galactosidase activity at the different pH values was determined as a percentage of the pH optimum. To obtain the pH profiles for crude β -galactosidase solutions the pH values versus relative activity were plotted.

2.4 Results and Discussion

2.4.1 Genomic DNA purification

The first step in the cloning process of the *lacZ* gene was the extraction of gDNA from *E. coli* W3110 to act as a template for the subsequent gene amplification (section 2.3.3).

After isolation, the purified gDNA was visualised on a 0.8 % (w/v) agarose gel (Figure 2.2) and gDNA was visible for all samples as a single band greater than 10 kb in size. The gDNA of *E. coli* K12 is over 4.6 million base pairs [Blattner *et al.* 1997]. The single band indicated that the gDNA isolated is of high quality and if shearing occurred then the fractions are larger than 10 kb. Since the *lacZ* gene is just over 3 kb intact copies of the gene should be present in the gDNA preparation.

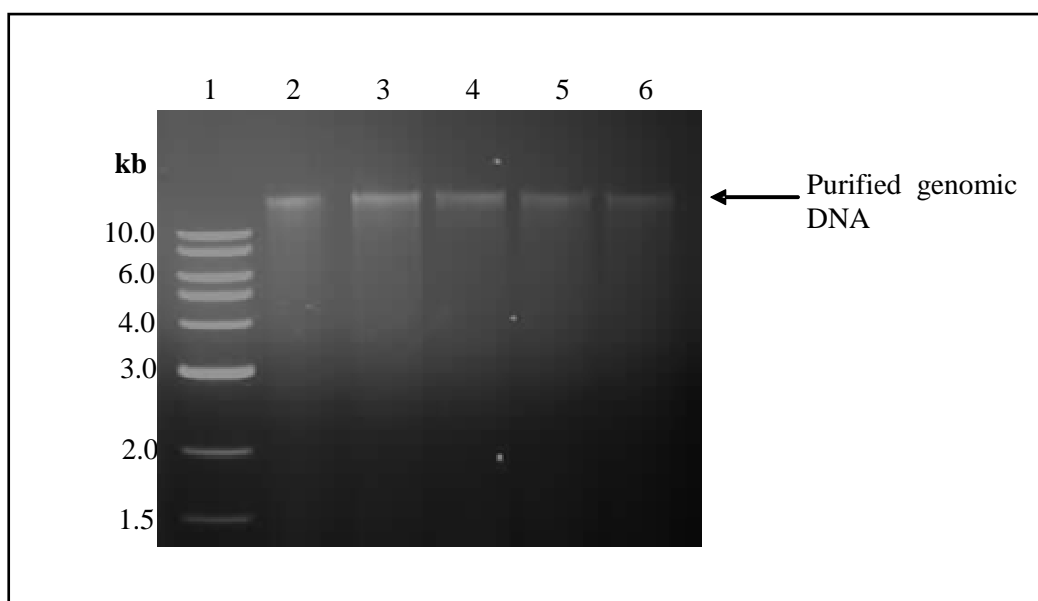


Figure 2.2 Electrophoresis analysis of purified genomic DNA from five overnight liquid cultures of *E. coli* W3110

Lane 1: DNA Ladder; Lanes 2-6: purified gDNA analysed on a 0.8 % (w/v) agarose gel stained for DNA with ethidium bromide.

Absorbance analysis of the gDNA preparation at a wavelength of 230 nm indicated that there was gDNA present but carryover of organic compounds or chaotropic salts from the spin columns of the DNA extraction kit during gDNA isolation occurred. It may also contain some contaminating proteins as peptide bonds absorb at 230 nm [Gallagher and Desjardins 2008]. Highly purified gDNA has a 260:280 nm ratio of

1.8-1.9 and contaminants (e.g. proteins) that absorb at 280 nm will lower the ratio. The ratio of absorbencies of 260 nm over 280 nm obtained was 1.71 which indicates that gDNA was purified but it also shows that there is a low level of contaminants present [Sambrook and Russell 2001; Gallagher and Desjardins 2008]. As the concentration of contaminants was minimal, the isolated gDNA was considered to be sufficiently pure and no difficulties were observed during the following cloning steps.

2.4.2 Amplification of *lacZ* gene using polymerase chain reaction

After successful gDNA isolation, polymerase chain reaction (PCR) was carried out as outlined in section 2.3.4 using the designed primers. The PCR cycle setup required optimisation for successful amplification of the *lacZ* gene. It was established that an annealing temperature of 58 °C while using ExTaq polymerase resulted in the highest product concentration and no impurities formed due to unspecified annealing. The ExTaq polymerase was employed not only due to its higher amplification yield but also due to its 3' to 5' exonuclease activity which ensures generation of longer, more accurate PCR products [White 1993].

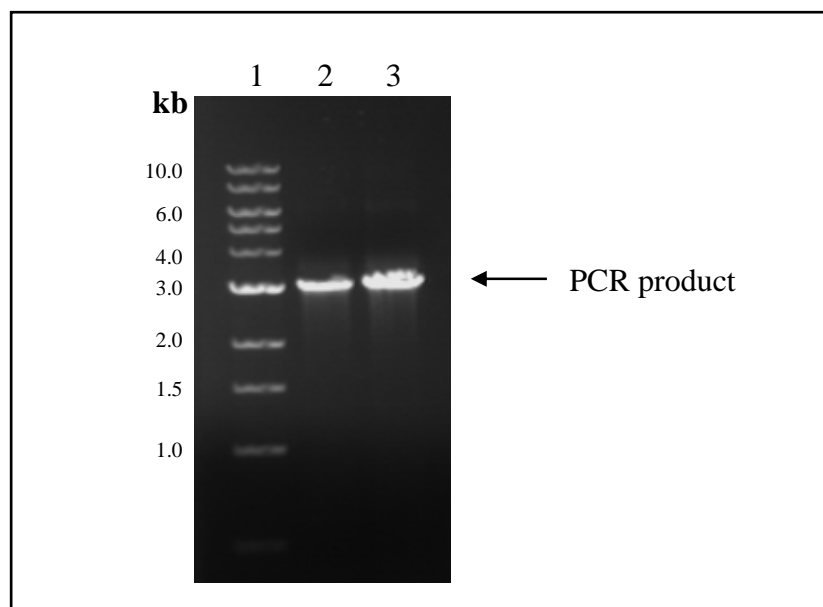


Figure 2.3 Electrophoretic analysis of PCR amplifications using optimised conditions for *lacZ* gene

Lane 1: DNA Ladder, Lanes 2 & 3: PCR product using template DNA from E. coli W3110 at optimal PCR conditions. The results are visualised on 0.8 % (w/v) agarose gel stained with ethidium bromide.

As seen in Figure 2.3 amplification resulted in a product between 3000 bp and 4000 bp which complies with the calculated length for the *lacZ* gene plus primers of 3093 bp. The PCR product was termed *lacZ*-HT indicating that after ligation into the pProEx-HTb vector the protein will be expressed with an N-terminal His₆-tag.

2.4.3 Purification and restriction of expression vector

The expression vector pProEx-HTb was propagated in *E. coli* DH5 and purified as per section 2.3.3. The isolated plasmid was subsequently digested using the restriction enzymes BamH1 and HindIII according to section 2.3.5. The results, illustrated in Figure 2.4, indicate successful purification and restriction of the pProEx-HTb. The plasmid was isolated in three forms: supercoiled, which migrated the furthest due to its condensed size, linear and open-circular (nicked) [Hardy 1986] and after the restriction digest only the linear form remained, identifying a length of just under 5 kb. The plasmid after digestion with the selected restriction enzymes (BamH1 and HindIII) is 4689 bp. The double digested plasmid, i.e. digested with both enzymes simultaneously, was used for the subsequent steps.

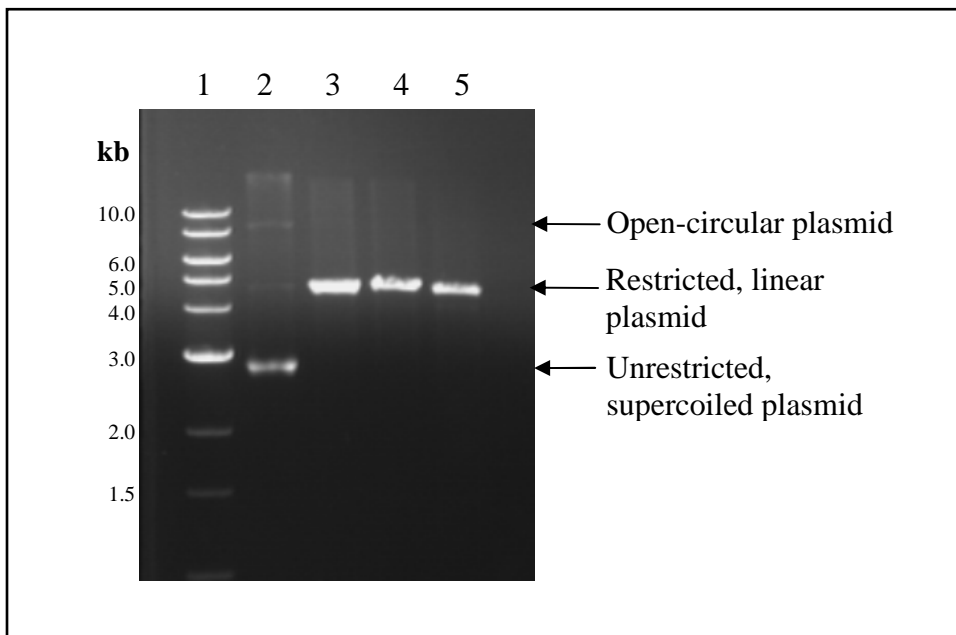


Figure 2.4 A 0.8 % (w/v) agarose gel depicting isolated pProEx-HTb and plasmid restriction results

Lane 1: DNA Ladder; Lane 2: purified pProEx-HTb before restriction digest; Lane 3: isolated pProEx-HTb restricted with BamH1; Lane 4: isolated pProEx-HTb restricted with HindIII; Lane 5: isolated pProEx-HTb double digested using both restriction enzymes (BamH1 and HindIII)

2.4.4 Purification of PCR product and subsequent restriction

PCR amplicons were pooled to obtain a sufficient concentration (0.5-1 g) and afterwards purified as outlined in section 2.3.3. The isolated PCR products were digested using the restriction enzymes BamH1 and HindIII as described in section 2.3.5 followed by further purification to remove the restriction enzymes. The results are shown in Figure 2.5 and indicate successful purification and restriction of the PCR products. The band of the purified gene indicates an appropriate size of approximately 3000 bp.

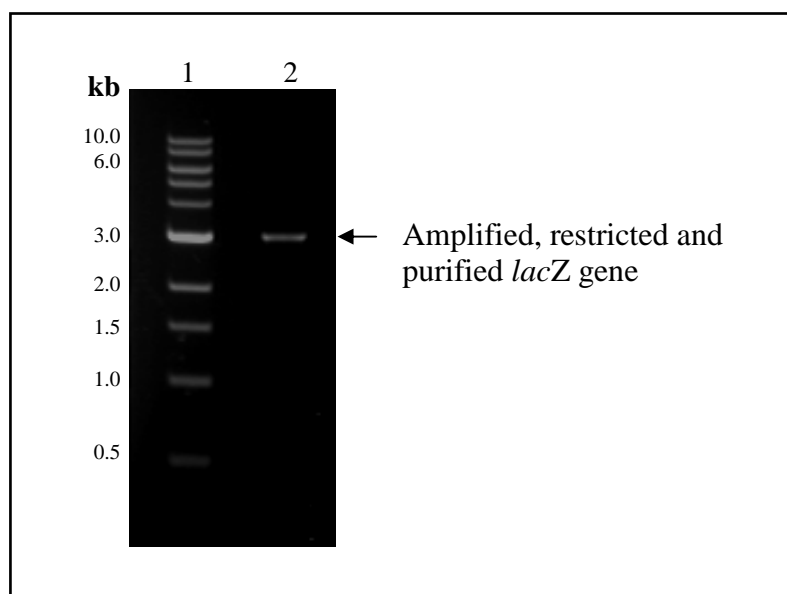


Figure 2.5 Electrophoretic analysis of restricted amplicon using the restriction enzymes BamH1 and HindIII

Lane 1: DNA Ladder; Lane 2: amplified lacZ gene restricted with BamH1 and HindIII. The restricted and purified gene was visualised on a 0.8 % (w/v) agarose gel stained with ethidium bromide.

2.4.5 Ligation of PCR product with plasmid and transformation into *E. coli* competent cells

Ligation of *lacZ*-HT into expression vector pProEx-HTb was carried out as outlined in section 2.3.5 which was followed by transformation into competent cells as described in section 2.3.6.

E. coli DH5 was employed in this study as the propagation strain for plasmids with and without *lacZ*-HT insert. This *E. coli* strain is recombination deficient (*recA1*)

and endonuclease A deficient (*endA*) which results in reduced occurrence of unwanted recombination in cloned DNA and allows cleaner preparation of DNA due to elimination of non-specific digestion by Endonuclease I [Grant *et al.* 1990]. After the ligation mixture was transformed into competent *E. coli* cells, the clones were grown on selective LB agar supplemented with ampicillin ensuring growth of strains carrying the plasmid only. Plasmid DNA was extracted to confirm the presence of the insert.

Eight randomly selected colonies, termed 1-8, were chosen for plasmid insert analysis. The plasmids of the transformants were purified according to section 2.3.3. The results of the plasmid isolation, illustrated in Figure 2.6, clearly indicate supercoiled plasmids of different sizes. The restricted pProEx-HTb plasmid consists of 4689 bp and the theoretical size of plasmids carrying the *lacZ*-HT gene is 7770 bp.

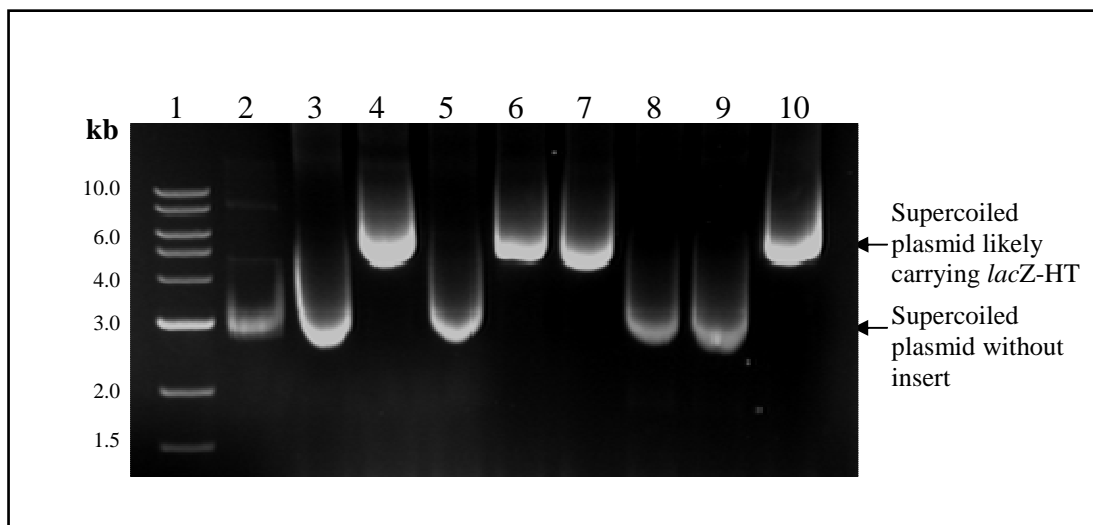


Figure 2.6 Electrophoretic analysis of purified plasmid DNA from *E. coli* DH5 transformants

Lane 1: DNA Ladder; Lane 2: purified pProEx-HTb (negative control for transformation); Lanes 3- 10: purified plasmid DNA from clones 1-8. The plasmid DNA was analysed on 0.8 % (w/v) agarose gel stained with ethidium bromide.

The significant difference in base pairs allows the distinction between plasmids with and without the *lacZ* insert. Transformants 1, 3, 6 and 7 (see Figure 2.6 Lanes 3, 5, 8 and 9) contained plasmids which run similar to the negative control, i.e. supercoiled plasmid known not to carry an insert. Clones 2, 4, 5 and 8 (Figure 2.6 Lanes 4, 6, 7

and 10) are likely to carry a pProEx-HTb with insert as they migrated slower on the same agarose gel, which is a sign of increased size, i.e. increased number of base pairs [Sambrook and Russell 2001].

The isolated plasmids from clones 2, 4, 5 and 8 were further investigated using restriction enzyme digestion as outlined in section 2.3.5. The results of the restriction analysis (Figure 2.7) show that the selected transformants had plasmids with an insert which is of the anticipated size of the β -galactosidase gene and is therefore likely to be *lacZ*-HT.

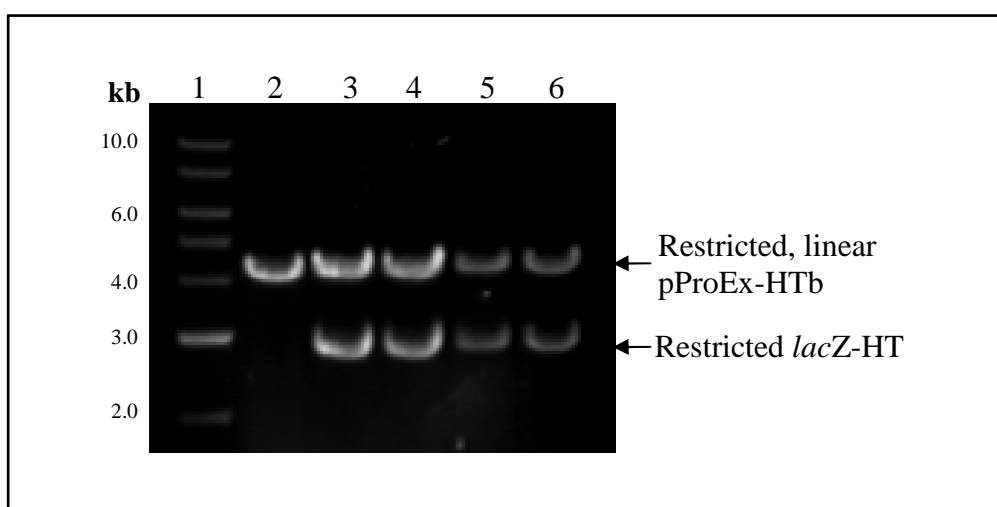


Figure 2.7 Purified and restricted plasmid DNA of *E. coli* DH5 transformants with pProEx-HTb carrying *lacZ*-HT insert visualised on 0.8 % (w/v) agarose gel

Lane 1: DNA Ladder; Lane 2: purified and restricted pProEx-HTb (negative control for transformation, i.e. no insert); Lanes 3-6: purified plasmid DNA of clones 2, 4, 5 and 8 restricted with BamHI and HindIII.

2.4.6 Sequencing analysis and homology modelling

Purified pProEx-HTb carrying the *lacZ*-HT gene from clone 2 was sent to MWG Biotech for sequencing analysis. The resultant sequence of *lacZ*-HT was aligned with the published sequence (accession number AP_000996.1) using ClustalW. The alignment produced a homology score of 99 and showed a point mutation for the recombinant *lacZ*-HT sequence at nucleotide 2274. Figure 2.8 is an excerpt of the ClustalW alignment showing the single point mutation highlighted in red. The full alignment of the recombinant *lacZ*-HT nucleotide sequence from the MWG analysis and the published gene sequence of *lacZ* from W3110 is illustrated in Appendix D. The single base substitution of adenine to a guanine was most likely introduced

during the cloning process. The *lacZ* sequence (accession number AP_000996.1) and the recombinant *lacZ*-HT sequence were translated using the translate tool from the ExPASy Proteomics server. The resulting amino acid sequences were aligned using the ClustalW program and showed that there is 100 % identity between both sequences (see Appendix E). The point mutation occurred at the third nucleotide of the codon encoding the 758 amino acid in the protein. This alteration changed the codon from CAA to CAG, however, due to the degeneracy of the genetic code both codons translate into the same amino acid (glutamine) [Nelson and Cox 2005].

```

W3110_lacZ      CTGACCACCAGCGAAATGGATTTTTCATCGAGCTGGGTAATAAGCGTTGGCAATTAAAC 2280
lacZ-HT        CTGACCACCAGCGAAATGGATTTTTCATCGAGCTGGGTAATAAGCGTTGGCAATTAAAC 2280
*****

```

Figure 2.8 Excerpt of ClustalW alignment of recombinant *lacZ*-HT nucleotide sequence from the MWG analysis and published gene sequence of *lacZ* from W3110 (accession number AP_000996.1) showing the single base mutation (red)

2.4.7 Expression of recombinant β -galactosidase

The pProEx-HTb plasmid carrying the *lacZ*-HT gene was transformed into an alternative expression host to avoid co-purification of the native β -galactosidase from *E. coli* DH5 . The strain selected for recombinant protein expression is *E. coli* MC1061 due to its (lac)X74 genotype resulting in a complete deletion of the *lac* operon. Plasmids confirmed to carry the *lacZ*-HT gene (clone 2) were purified and transformed into competent *E. coli* MC1061 cells according to sections 2.3.3 and 2.3.6, respectively.

Expression of recombinant β -galactosidase in *E. coli* MC1061 was verified by carrying out a small-scale induction as outlined in section 2.3.7. The resulting crude extracts were analysed on 8 % SDS-PAGE (section 2.3.8.1) which showed that the clone produced a protein with a distinct overexpressed band of approximately 116 kDa, the molecular mass of the *E. coli* β -galactosidase protein, as illustrated in Figure 2.9.

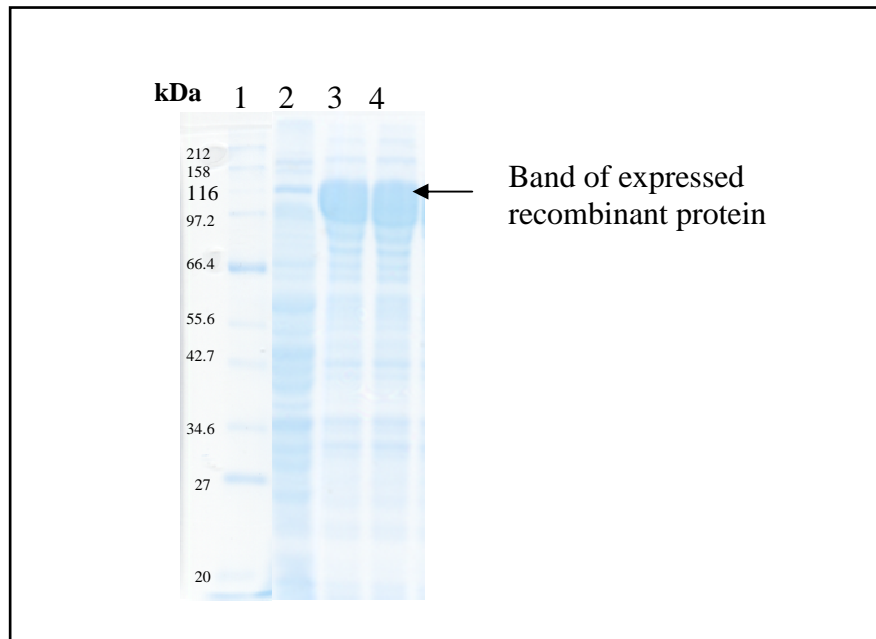


Figure 2.9 SDS-PAGE analysis of recombinant β -galactosidase expressed in *E. coli* MC1061

*Lane 1: Broad range molecular mass marker (New England Biolabs); total protein crude extract of uninduced control (*E. coli* MC1061 pProEx-HTb_lacZ-HT); Lane 3: total protein crude extract of culture induced with IPTG (final concentration 0.25 mM); Lane 4: total protein crude extract of culture induced with IPTG (final concentration 0.5 mM).*

Further expression analysis during optimisation studies indicated that the majority of the recombinant protein was expressed as non-soluble aggregates. Protein bands with an estimated molecular mass of 116 kDa could be observed in the soluble and non-soluble crude solutions of expression cultures with an increased protein concentration in the non-soluble fraction. These results are illustrated in Figure 2.10. The uninduced control also shows bands at 116 kDa in both protein fractions which illustrates that the *trc* promoter in the plasmid allows a certain amount of basal or “leaky” expression, i.e. the recombinant protein is produced without the presence of the inducer (IPTG) [Ham *et al.* 2006]. No protein band with that molecular mass could be observed in crude extracts of *E. coli* MC1061 carrying pProEx-HTb plasmid without an insert.

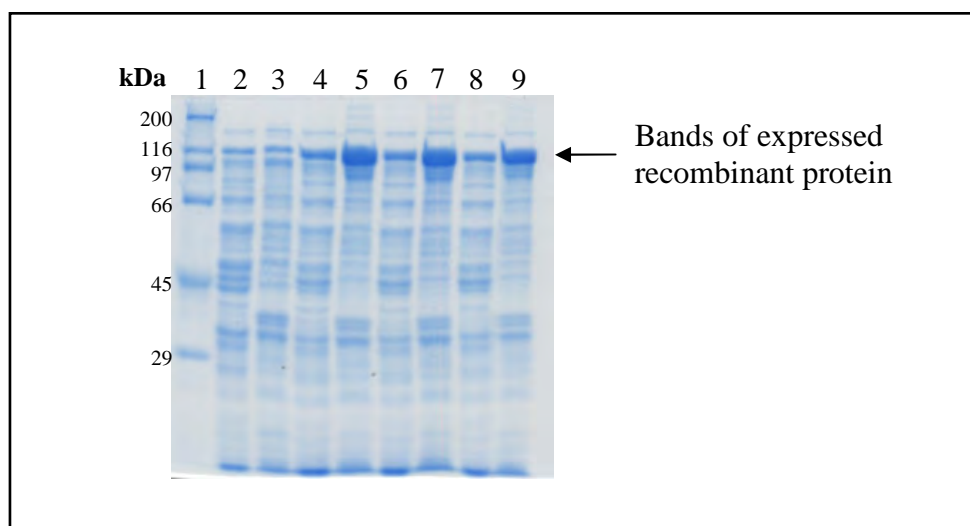


Figure 2.10 Electrophoretic results of IPTG concentration optimisation studies for recombinant β -galactosidase expression

Lane 1: Molecular mass marker (Sigma-Aldrich); Lane 2: soluble fraction of uninduced control; Lane 3: non-soluble fraction of uninduced control; Lane 4: soluble fraction of induced with final IPTG concentration of 0.05 mM; Lane 5: non-soluble fraction of induced with final IPTG concentration of 0.05 mM; Lane 6: soluble fraction of induced with final IPTG concentration of 0.08 mM; Lane 7: non-soluble fraction of induced with final IPTG concentration of 0.08 mM; Lane 8: soluble fraction of induced with final IPTG concentration of 0.1 mM; Lane 9: non-soluble fraction of induced with final IPTG concentration of 0.1 mM. The crude extracts were analysed on an 8% SDS-PAGE stained for protein with EZBlue.

During expression of recombinant proteins in *E. coli* the formation of soluble, active protein as well as insoluble, inactive protein can be observed regularly. The non-soluble protein aggregates, or inclusion bodies, are misfolded versions of the recombinant protein product and accumulate due to irregular or incomplete folding processes. Inclusion bodies form during recombinant protein translation due to the high levels of expression to concentrations found rarely in nature which overwhelms the endogenous protein folding machinery [Mitraki and King 1989; Wilkinson and Harrison 1991; Ventura and Villaverde 2006]. In the biopharmaceutical industry, the formation of inclusion bodies is undesirable as these insoluble protein aggregates are non-functional. While additional processing steps can be employed to facilitate denaturation and refolding of the protein, renaturation to functional protein is not guaranteed and even at optimised conditions can give relatively low yields [Lilie *et al.* 1998]. These added steps can also increase the downstream processing costs significantly.

It has been shown that reducing the protein synthesis rate by lowering the inducer concentration or the expression temperature can reduce the extent of inclusion body formation [Wilkinson and Harrison 1991; Krause *et al.* 2010]. Therefore, expression of recombinant β -galactosidase was induced at 25 °C after the *E. coli* MC1061 expression clones were grown at 37 °C to reach a sufficient cell density. Figure 2.11 shows the soluble and non-soluble crude extract fractions of recombinant *E. coli* MC1061 induced with different concentrations of IPTG after expression at 25 °C for 16 h. For all induced cultures there is a band of overexpressed protein visible with an approximate molecular mass of 116 kDa. The culture induced with 0.08 mM IPTG appeared to produce highest ratio of soluble protein as determined in Figure 2.11 and assessed via standard activity assays (section 2.3.9). This was identified as the optimal IPTG induction concentration for the expression of recombinant β -galactosidase at 25 °C. Future production of the recombinant protein employed these conditions, however, expression time was reduced to 6 h, as it results in sufficient quantities of product.

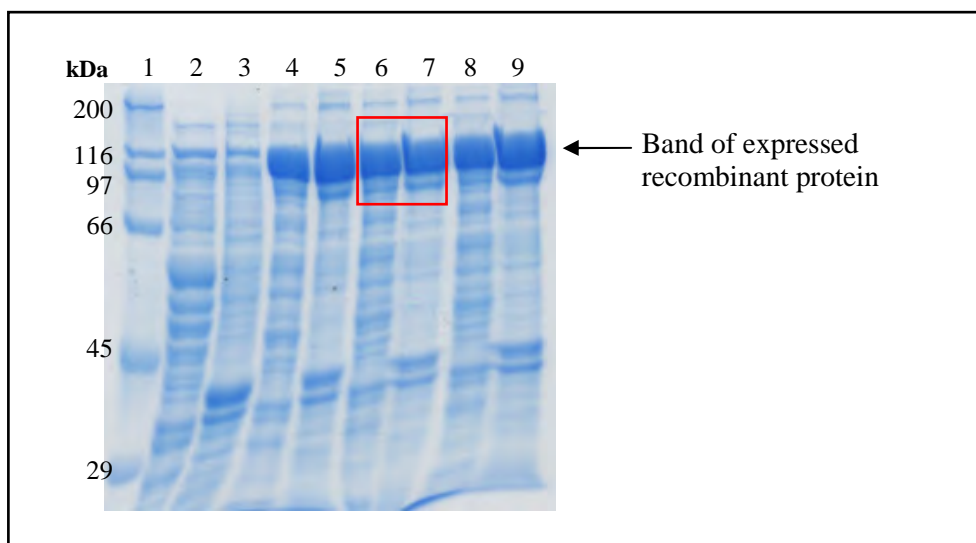


Figure 2.11 SDS-PAGE analysis of the crude extracts obtained during IPTG optimisation studies at 25 °C

The crude extracts were analysed on an 8 % SDS-PAGE stained for protein. Lane 1: Molecular mass marker (Sigma-Aldrich); Lane 2: soluble fraction of uninduced control; Lane 3: non-soluble fraction of uninduced control; Lane 4: soluble fraction of induced with final IPTG concentration of 0.05 mM; Lane 5: non-soluble fraction of induced with final IPTG concentration of 0.05 mM; Lane 6: soluble fraction of induced with final IPTG concentration of 0.08 mM; Lane 7: non-soluble fraction of induced with final IPTG concentration of 0.08 mM; Lane 8: soluble fraction of induced with final IPTG concentration of 0.1 mM; Lane 9: non-soluble fraction of induced with final IPTG concentration of 0.1 mM. Highlighted with a red box is the optimal IPTG concentration, which was used for future expression of the recombinant protein, as it achieves the highest ratio of soluble protein.

2.4.8 Native PAGE and Western blot of recombinant β -galactosidase

Native gels with the crude extract from *E. coli* MC1061 pProEx-HTb_ *lacZ*-HT after expression of the recombinant protein were run as described in section 2.3.8.2. Figure 2.12 depicts the native gels stained for protein (Lane 1) and for activity (Lane 2). The zymogram (activity gel) confirmed β -galactosidase activity in the crude extract of the recombinant MC1061 strain. As this *E. coli* strain is a *lacZ* deletion mutant the β -galactosidase activity is due to the recombinant expression of *lacZ*-HT insert carried by the pProEx-HTb plasmid. The protein and the activity gel both show an additional band with a higher molecular mass than the recombinant β -galactosidase. This band most likely represents low levels of a higher order oligomeric version of the recombinant protein [Juers *et al.* 2000].

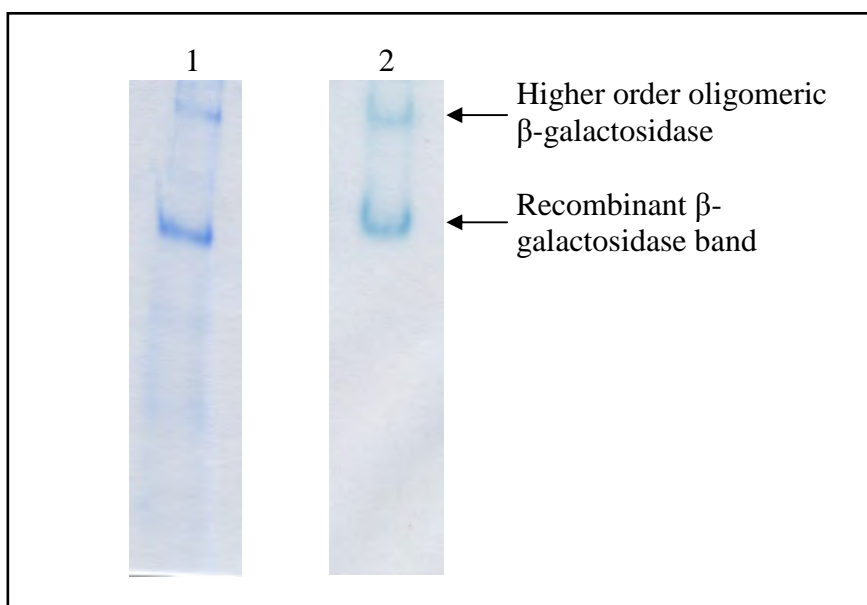


Figure 2.12 Native gels showing crude extract from *E. coli* MC1061 pProEx-HTb_ *lacZ*-HT stained for protein and activity

Lane 1: Crude extract from E. coli MC1061 carrying pProEx-HTb_lacZ-HT run on 5.0 % native gel stained for protein; Lane 2: Crude extract from E. coli MC1061 carrying pProEx-HTb_lacZ-HT run on 5.0 % native gel stained for activity using x-gal.

Western blot analysis on the crude extract from *E. coli* MC1061 pProEx-HTb_ *lacZ*-HT after expression of the recombinant β -galactosidase at optimised conditions was carried out according to section 2.3.8.3. The plasmid pProEx-HTb facilitates the addition of an N-terminal polyhistidine tag to the protein during expression. The antibody selected (Anti-polyHistidine Peroxidase conjugate) binds to the His₆-tag

and thereby confirms the presence of the recombinant product as well as the presence of the tag. Figure 2.13 shows the results of the Western blot which verify that a recombinant protein with a His₆-tag is being expressed in the selected clone.

The additional band visual both on the SDS-PAGE and Western blot in Figure 2.13 with a molecular mass between 97 and 66 kDa is most likely a 90 kDa β -galactosidase degradation fragment. It is a frequently observed break-down product found in both native and engineered *E. coli* β -galactosidases [Kosinski and Bailey 1991; Corchero *et al.* 1996; Viaplana *et al.* 1997; Corchero and Villaverde 1999]. This intermediate arises from a proteolytic cleavage at the C-terminal of the polypeptide [McKnight and Fried 1981] as was shown by β -complementation studies. Consequently, a band could still be detected during Western blot analysis as the antibody binds to the His₆-tag which is N-terminally linked. The cleavage site is at a presumed trypsin-like target site between Arg611 and Thr612 [Viaplana *et al.* 1997] and a trypsin inhibitor was added to all buffers during this study to reduce the proteolytic cleavage of the recombinant product.

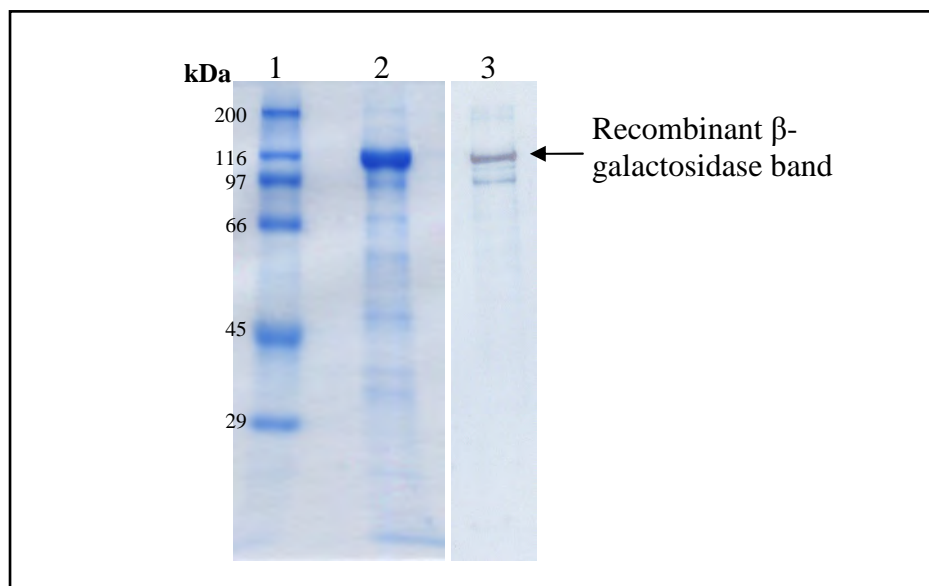


Figure 2.13 SDS-PAGE and Western blot analysis of recombinant β -galactosidase expressed in *E. coli* MC1061

*Lane 1: Molecular mass marker (Sigma-Aldrich); Lane 2: Crude extract from *E. coli* MC1061 carrying pProEx-HTb_lacZ-HT run on 8.0 % SDS-PAGE gel stained for protein; Lane 3: Crude extract from *E. coli* MC1061 carrying pProEx-HTb_lacZ-HT transferred to blot and stained for His₆-tag*

Genetic engineering enabled the production of a model *E. coli* strain expressing soluble and functional recombinant β -galactosidase with an N-terminal His₆-tag.

2.4.9 Comparison of native and recombinant β -galactosidase

The crude native (from *E. coli* W3110) and recombinant β -galactosidase enzymes were compared by determining their enzymatic activities at different pH and temperature conditions. These results allowed the construction of a pH versus relative activity profile and a temperature versus relative activity profile. This work was undertaken to validate the model system produced by showing a functional equivalency between the native and recombinant enzyme.

The pH activity profile for the native and recombinant β -galactosidase was generated as outlined in section 2.3.11 and is shown in Figure 2.14.

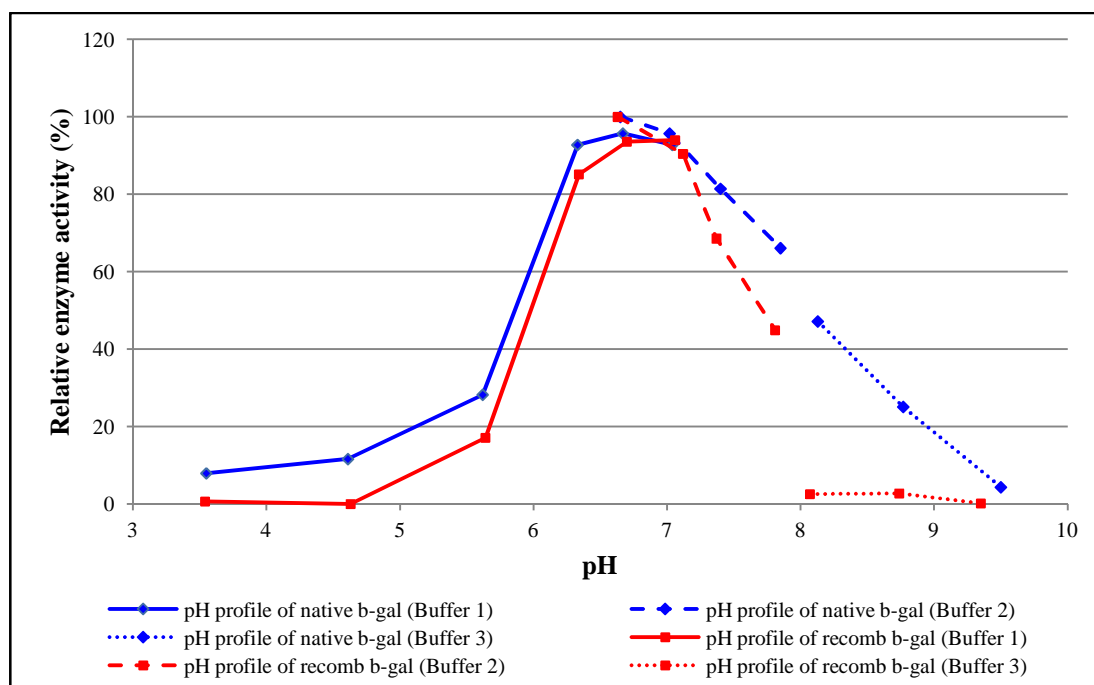


Figure 2.14 pH versus relative activity profile of the native and recombinant β -galactosidase enzyme

Buffer 1: 0.2 M citric acid- Na_2HPO_4 ; Buffer 2: 0.2 M sodium phosphate buffer; Buffer 3: 0.2 M boric acid- NaOH

The data is expressed as a percentage of maximum activity and each value represents mean \pm standard deviation ($n = 3$; for native enzyme 0.03 EU and for recombinant galactosidase 0.02 EU per reaction).

The results obtained indicate that native and recombinant enzymes exhibit a very comparable pH versus relative activity behaviour between pH 5.5 and pH 7.5. Both enzymes displayed an optimum activity at pH 6.5 which is similar to the recorded optimum of pH 7.0 for the hydrolysis for the synthetic substrate ONPG [Flickinger

and Drew 1999]. However, at a pH of 8.0 or higher (boric acid-NaOH buffer) the recombinant β -galactosidase appears to lose all enzymatic activity while the native enzyme still retains some biological function up to pH 9.0.

The temperature versus relative activity profile was obtained as described in section 2.3.10 and the results are illustrated in Figure 2.15.

The temperature profiles for both enzymes are similar at lower temperatures (≤ 50 °C) but also indicate that the native enzyme is more stable at higher temperatures. The maximum activity of native β -galactosidase from *E. coli* W3110 maximum activity is between 50 and 55 °C and it still retains about 40 % of its relative activity at 60 °C. The recombinant β -galactosidase has its maximum activity at 50 °C and loses more than half its relative activity at 55 °C.

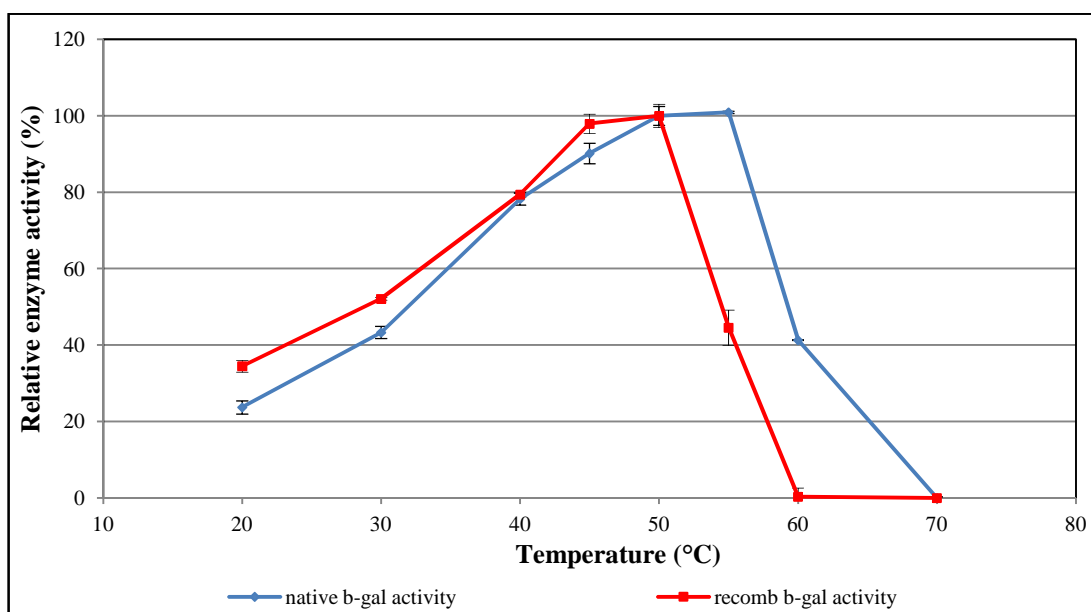


Figure 2.15 Temperature versus relative activity profile for the native and recombinant β -galactosidase enzyme

The data is presented as a percentage of maximum activity and each value is shown as mean \pm standard deviation ($n = 3$; for native enzyme and recombinant β -galactosidase 0.03 EU per reaction).

These results indicate that the native enzyme is more stable at higher pH and temperature levels than the recombinantly produced enzyme. Both proteins have the same amino acid sequence with the exception of the His₆-tag and its cleavage site at

the N-terminal end of the recombinant β -galactosidase which may explain the difference observed [Waugh 2005].

Additionally, the native protein may be more stable due to stabilisers present in the crude cellular extract. For example, mercaptoethanol has been shown to increase the heat stability of the *E. coli* β -galactosidase enzyme enabling it to retain 50 % of biological function after 40 min at 55 °C whereas all activity was lost after 10 min without mercaptoethanol [Craven *et al.* 1965]. The recombinant protein, due to its high concentration in the cell extract, requires more extensive dilutions prior to testing (1 in 800-1,000) compared to the native enzyme. Any possible stabilisers in the crude extract may be too dilute to be effective.

In conclusion, both crude enzyme solutions exhibit the same pH optimum of pH 6.5 and maximum activity at approximately 50 °C which shows that there is very substantial functional equivalency between the native and recombinant β -galactosidase.

2.5 Conclusion

As part of this study it was necessary to generate an *Escherichia coli* expression system capable of producing a recombinant protein. Standard molecular engineering techniques were employed to introduce a bacterial β -galactosidase coding sequence (*lacZ*) into *E. coli* MC1061. The resulting model strain (*E. coli* MC1061 carrying the plasmid pProEx-HTb_ *lacZ*-HT) expressed a soluble, functional β -galactosidase enzyme with an N-terminal His₆-tag of the correct molecular mass. Expression studies were employed to maximise the production of soluble protein.

This *E. coli* model expression system can now be employed to investigate the potential minimisation of phosphorus during fermentation, i.e. strain propagation and protein expression (upstream processing), as well as product purification (downstream processing).

**Chapter Three – Assessment of reducing the phosphorus
content in fermentation media for recombinant protein
production in *E. coli***

3.1 Introduction

In the biotechnology sector the main goal during upstream processing is the production of high cell and hence product yields. Media ingredients are relatively inexpensive and are often added in excess of actual growth and protein expression requirements to ensure high cell density cultures. The environmental impacts of unutilised media constituents are traditionally given little consideration during process development [Cliffe *et al.* 2010].

Phosphorus (P) is an essential nutrient in media for *E. coli* growth and recombinant product formation as it is required by the cell for the production of nucleic acids, phospholipids and cofactors. In addition, inorganic P is involved in the transport of cellular energy in the form of ATP [Prescott *et al.* 2002; Madigan *et al.* 2009]. This makes P an essential core nutrient for *E. coli* fermentation and recombinant protein production, however, P-containing compounds can have detrimental effects upon the environment in particular in terms of nutrient enrichment. While natural eutrophication is a slow process occurring over a long period of time, the presence of human activities has hastened the process by elevating the increase of nutrient inputs into water bodies [Khan and Ansari 2005]. P is frequently regarded as the growth-limiting nutrient in natural waters with little or no pollution and excessive P concentration is considered the most common cause of eutrophication in freshwater, coastal and estuarine systems [Morton and Edwards 2005]. Fermentation wastes from the biotechnology sector include large amounts P-containing materials [Cliffe *et al.* 2010], which along with nitrogen can accelerate the process of eutrophication.

The element of research described in this chapter aimed to investigate the P concentration in media commonly employed in recombinant *Escherichia coli* fermentation as well as the determination of the potential of P minimisation without negatively affecting biomass or recombinant product yields. A reduction of the P content in fermentation media would be very beneficial in terms of the environmental impact of the subsequent wastes from this process. The microbial strain employed during this study is the model strain designed *E. coli* MC1061 producing recombinant His₆-tagged β -galactosidase as outlined in chapter two.

3.2 Materials

As described in section 2.2.

3.3 Experimental methods

3.3.1 Fermentation and cell lysis

Fermentation of the recombinant *E. coli* strain was carried out in 500 ml conical flasks with culture volumes of 100 ml using the optimised expression conditions (section 2.4.8). The media were inoculated with 1.0 ml of overnight *E. coli* precultures (grown in LB supplemented with ampicillin for 15 h at 37 °C, 250 rpm) diluted to an absorbance value of 1.5 at 600 nm with fresh LB media. All media employed are described in section 2.2.4 and were supplemented with ampicillin (100 g ml⁻¹) to ensure continued presence of pProEx-HTb_ *lacZ*-HT. Upon completion of expression samples for absorbance at 600 nm (section 3.3.2), dry cell weight determination (sections 3.3.3) and enzyme activity (section 3.3.5) were taken or the cells were harvested and lysed as outlined in section 2.3.1.

3.3.2 Absorbance values at 600 nm

The cell growth was monitored by measuring the optical density at 600 nm (OD₆₀₀) of 1.0 ml culture sample using the UV-visible spectrophotometer (Shimadzu UVmini 1240, Shimadzu Scientific Instruments, Torrance, USA). The blank samples used during this procedure are the fresh media (not inoculated) employed during the study.

3.3.3 Determination of dry cell weight (DCW)

Dry cell weight (DCW) was determined via an adapted method described by Suárez *et al.* [1998]. 25 ml of culture samples were transferred into pre-dried and pre-weighed universal tubes which were subsequently centrifuged at 7,500 rpm, 4 °C for 25 min. The resulting cell pellets were washed with 10 ml sterile water and again centrifuged at 7,500 rpm, 4 °C for 15 min. The pellets were then dried in a 100 °C oven for 24 h, allowed to cool and the grams of DCW per litre of culture volume (g(DCW) l⁻¹) was determined.

3.3.4 Growth of *E. coli* on selected media

Overnight precultures of the *E. coli* model strain diluted to an absorbance value of 1.5 at 600 nm with fresh LB media were used to inoculate 100 ml of each culture in 500 ml conical flasks. All flasks were incubated at 37 °C, 200 rpm. The growth of the model strain in the various media was observed every hour via absorbance at 600 nm (section 3.3.2) for a total of 14 h from the point of inoculation. Microbial growth curves were constructed with absorbance at 600 nm versus time (in h) for the model strain in each media.

3.3.5 β -Galactosidase assay and kinetic analysis

The enzyme activity was assayed as outlined in section 2.3.9. Kinetic data was obtained by assaying the four soluble crude extract samples of the model strain expressed in the various media. Determination of the kinetic constants was achieved by monitoring the rate of ONPG hydrolysis over time at substrate concentrations ranging from 0.1-2.0 mM in triplicate. Lineweaver-Burk plots and initial velocity against substrate concentration plots were constructed allowing determination of the kinetic parameters by linear regression [Henderson 1993].

3.3.6 Bradford assay

The Bradford assay was carried out according to the procedure described by Bradford [1976]. The Bradford reagent consisted of 100 mg of Commassie Brilliant Blue-G250, 50.0 ml of 95 % (v/v) ethanol, 100 ml 85 % (v/v) phosphoric acid and distilled water diluted to 1 l. The reagent was filtered through Whatman No 1 filter paper and stored in a dark bottle at 2-8 °C which kept it stable for several months.

The assay procedure involved mixing 1.0 ml of Bradford reagent with 100 μ l of protein sample and incubating it at room temperature for 2 min prior to measuring its absorbance at 595 nm. The absorbance values were converted to protein concentration using a standard calibration curve constructed with bovine serum albumin (BSA) ranging from 10 to 200 μ g(BSA) ml⁻¹.

3.3.7 Determination of recombinant protein concentration from SDS-PAGE

The concentration of recombinant protein produced in *E. coli* MC1061 pProEx-HTb_ *lacZ*-HT was determined via SDS-PAGE analysis according to Crespo *et al.* [2008] with some modifications. Cell pellets were resuspended in 500 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄, pH 7.0, 10 mM MgCl₂, 10 mM β-mercaptoethanol, 50 mg l⁻¹ soybean trypsin inhibitor to a cell concentration of 20 mg(DCW) ml⁻¹. The cell suspensions were sonicated as described in section 2.3.1 and fractionated by centrifugation at 10,000 rpm, 4 °C for 10 min. The soluble fractions (supernatant), non-soluble fractions (pellet) and total protein fractions (solution prior to centrifugation) were mixed with SDS-PAGE loading buffer and run on 7 % SDS-PAGE gels according to section 2.3.8.1. A standard curve of commercial β-galactosidase (Sigma-Aldrich; G8511) concentration versus density of the resulting bands on an SDS-PAGE gel was constructed. β-Galactosidase standard solutions, ranging from 0-2.0 μg of enzyme per lane, were prepared by diluting a stock solution of 2.0 mg(β-gal) ml⁻¹ in 500 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄, pH 7.0, 10 mM MgCl₂, 10 mM β-mercaptoethanol, 50 mg l⁻¹ soybean trypsin inhibitor. The standard solutions were also run on a 7 % SDS-PAGE gel and both standard and sample gels were stained according to the procedure outlined in Table 3.1.

Table 3.1 Staining procedure for SDS-PAGE gels used for determination of recombinant protein concentration

Step	Time	Temperature
Washing (distilled water)	3x5 min	Room temperature
Fixing (10 % (v/v) acetic acid, 50 % (v/v) methanol)	15 min	22 °C
Washing (dist. water)	15 min	22 °C
EZBlue staining solution	60 min	22 °C
Destaining (distilled water)	16 h (until background clear)	22 °C

The gels were analysed using the UVP biospectrum 410 imaging system and the Vision Works[®] LS software. The concentration of recombinant β-galactosidase was

determined from the density of the resulting β -galactosidase band and correlating it to the standard curve.

3.3.8 Total phosphorus analysis

Total phosphorus (TP) includes all dissolved and particulate phosphorus that may be found in all unfiltered samples [Maher and Woo 1998]. The TP concentration was determined according to approved standard methods 4500-P [Eaton and Franson 1995] which entails a two-step procedure: the first step is the conversion of phosphorus forms to dissolved orthophosphate which was carried out using the 4500-P B persulphate digestion method. The second step is the colorimetric determination according to the 4500-P D stannous chloride method. For the digestion a 50 ml sample was mixed with 1.0 ml of sulphuric acid solution (5 N H_2SO_4) and 0.4 g of ammonium persulphate ($(\text{NH}_4)_2\text{S}_2\text{O}_8$). The solution was boiled until a reduction to approximately 10 ml was achieved.

The digested sample was diluted to 100 ml with distilled water and 4 ml of molybdate reagent 1 (ammonium molybdate ($(\text{NH}_4)_6\text{Mo}_7\text{O}_{24}\cdot\text{H}_2\text{O}$) (25 g) was dissolved in 175 ml distilled water while 280 ml concentrated sulphuric acid was added to 400 ml distilled water, both solutions were combined and diluted to 1 l) and 0.5 ml stannous chloride reagent 1 (2.5 g stannous chloride ($\text{SnCl}_2\cdot 2\text{H}_2\text{O}$) was dissolved in 100 ml glycerol) were added and the solutions were mixed. After 10 min but before 12 min the absorbance of the samples was determined spectrophotometrically at 690 nm. The absorbance values were compared with the calibration curve using distilled water as a blank.

A standard curve of absorbance at 690 nm versus P concentration ($\text{mg(P)}\text{ l}^{-1}$) was constructed to quantify the amount of TP in the samples. Standard phosphorus solutions, ranging from 0.02 to 2.0 $\text{mg(P)}\text{ l}^{-1}$, were prepared in triplicate by diluting a 50 $\text{mg(P)}\text{ l}^{-1}$ stock solution of potassium phosphate (KH_2PO_4) in distilled water. The standard solutions were digested using the 4500-P B persulphate digestion method and the absorbance at 690 nm was determined according to the 4500-P D stannous chloride method [Eaton and Franson 1995].

3.3.9 Statistical analysis

The statistical analysis was carried out using Minitab[®] Statistical software, Release 15 for Windows. Where appropriate, the data generated was analysed by one-way analysis of variance (ANOVA) as it determines if there is a statistically significant difference between level means for each value. The statistical significance of results was evaluated through the probability value (p -value; < 0.05) which indicates if the level means are significantly different from each other [Brase and Brase 2009].

3.4 Results and Discussion

3.4.1 Media selection

Information on large-scale commercial production of approved biopharmaceuticals was analysed but little scientific data on which media or which type of media (complex, chemically defined or semi-defined) was available as this information is considered highly confidential by the producer companies. For the purpose of this study Luria-Bertani (LB) broth and Terrific broth (TB) were chosen as representatives for complex media as they appear to be the most frequently employed media in the general scientific literature for the production of recombinant proteins in *E. coli* [Aristidou *et al.* 1999; Sambrook and Russell 2001; Atlas 2004; Fu *et al.* 2006]. Similar media with complex carbon and nitrogen sources are commonly utilised to support high productivity in industrial-scale fermentation systems [Davies and Demain 1999]. In addition, direct communication with a senior production scientist within Schering-Plough (Brinny, Co. Cork) confirmed that *E. coli*-based biopharmaceutical production methods do employ substantially similar media to those described [personal communication]. Minimal media 9 (M9) or a variation thereof is one of the most regularly used chemically defined media for *E. coli* fermentation [Sambrook and Russell 2001]. It was included in this study, as the use of defined media components is likely to be a trend evident in the manufacture of recombinant therapeutic proteins. This type of media often results in more consistent titres, allows easier process control and monitoring and simplifies downstream processing. Additionally, the absence of animal-derived raw materials reduces extensive testing of the raw materials and makes this media a safer manufacturing option [Zhang and Greasham 1999; Tripathi *et al.* 2009a]. Yeast extract (YE) was added to the M9 media to improve biomass yields.

3.4.2 Phosphorus analysis of all fermentation media

The total phosphorus (TP) of individual (phosphorus-containing) media components was determined according to section 3.3.8 and the results of this analysis are illustrated in Table 3.2. The TP content varied significantly between the three media, with TB containing the highest concentration at 2.74 g(P) l⁻¹, followed by M9/YE with 1.77 g(P) l⁻¹. To place these findings into an industrially relevant context, these

values would represent 27.40 and 17.70 kg of phosphorus per 10,000 l batch. Industrial-scale bioreactors vary in size from 1,000 l to 100,000 l with typical recombinant DNA-based biopharmaceuticals ranging from 10,000 l upwards [Junker 2004]. The exact fermentation capacity for any biopharmaceutical is dictated by market demand and such information is highly confidential. Using a batch size of 10,000 l is a realistic but likely conservative estimate for this study.

Table 3.2 The chemical composition of the three media studied (LB, TB, M9/YE) along with the total phosphorus (TP) concentrations experimentally determined (section 3.3.8) in the P-containing media constituents

	LB		TB		M9/YE	
	Component (g l ⁻¹)	TP (g(P) l ⁻¹)	Component (g l ⁻¹)	TP (g(P) l ⁻¹)	Component (g l ⁻¹)	TP (g(P) l ⁻¹)
Tryptone	10.00	0.08	12.00	0.08		
Yeast extract	5.00	0.06	24.00	0.25	2.00	0.03
Glucose					4.00	
Glycerol			4.00 ml			
NaCl	10.00				0.50	
KH ₂ PO ₄			12.54	1.92	3.00	0.58
K ₂ HPO ₄			2.31	0.50		
NaH ₂ PO ₄ ·7H ₂ O					12.80	1.15
NH ₄ Cl					1.00	
MgSO ₄					0.2	
CaCl ₂					0.01	
TP for each media (g(P) l⁻¹)	0.13		2.74		1.77	

It is also visible from the results that the majority of TP in all media originates from the phosphate salts added to TB and M9/YE. LB broth, which has no defined P sources added, has a much lower TP concentration of 0.13 g(P) l⁻¹ in comparison to the other media investigated. This media's P content is solely derived from the complex nutritional components: tryptone and yeast extract. These complex ingredients, added to all media, comprise of components some of which have incorporated phosphorus as well as other necessary nutrients required for cell growth and protein expression [Krause *et al.* 2010]. As there are sufficient/near sufficient

quantities of P available through the complex media constituents, as shown in LB, it may be possible to reduce the amount of P salts added to the other two media.

Spent media samples, i.e. media post growth of *E. coli* model strain and expression of the recombinant protein, were analysed for TP concentrations to further evaluate the potential for P minimisation during fermentation. The results, summarised in Table 3.3, indicate that a substantial amount of P remains unutilised in the spent media. The P uptake by the cells during fermentation in LB and TB is modest as the TP analysis of the spent fermentation media showed residual unutilised TP of 77 % and 78 %, respectively. The model strain grown and expressed in the semi-defined media utilises even less phosphorus with 87 % TP remaining post fermentation. While previous studies have shown that too low concentration of P impairs cell growth [Korz *et al.* 1995], these findings clearly indicate a potential for significant reductions in the amount of exogenous P added to the media employed for the growth of bacteria and expression of recombinant proteins.

Table 3.3 The utilisation of total phosphorus (TP) by the model strain during growth and expression of the recombinant protein

Media	TP in fresh media (g(P) l ⁻¹)	TP in spent media (g(P) l ⁻¹)	Mean P utilised from fresh media (%)	Mean P unutilised in spent media (%)
LB	0.13 ± 0.003	0.11 ± 0.004	23	77
TB	2.74 ± 0.14	2.13 ± 0.10	22	78
M9/YE	1.77 ± 0.16	1.54 ± 0.07	13	87

The data shows mean TP values ± standard deviation (n = 6)

For the phosphorus minimisation study the quantity of exogenous P (KH₂PO₄, K₂HPO₄, NaH₂PO₄·7H₂O) added to TB and M9/YE media was reduced by 50, 70, 90 and 100 %. The unaltered media are described as 0 % P reduced for both media with compositions as outlined in Table 3.2 and are used as the 100 % control throughout this study. As the only ingredients in LB media containing P are the tryptone and yeast extract, it was not possible to reduce the P input in that media without affecting the concentration of other nutrients.

TP concentrations for all media with the varying amounts of added P were determined experimentally (section 3.3.8) and the results are listed in Table 3.4. The majority of TP in TB and M9/YE is due to the addition of the P salts, therefore, reducing these salts will reduce the overall TP concentration of the media as seen in Table 3.4.

However, as there is some P provided through the complex ingredients, even the 100 % P reduced media contains some phosphorus. In the case of TB there is still more than double the amount of TP in the 100 % P reduced media compared to LB due to the high concentrations of tryptone and yeast extract. In industrial-scale fermentation, these findings translate to P reductions from 27.40 to 3.90 kg and 17.70 to 0.30 kg per 10,000 l batch for TB and M9/YE, respectively. This indicates a potential that the environmental load of fermentation media, i.e. upstream processing in biopharmaceutical production, can be significantly reduced by decreasing the amount of exogenous P added.

Table 3.4 Total phosphorus (TP) concentrations experimentally determined for all media employed in this study

% P reduced	TP (g(P) l ⁻¹)		
	LB	TB	M9/YE
0	0.13 ± 0.003	2.74 ± 0.14	1.77 ± 0.16
50		1.47 ± 0.12	0.93 ± 0.15
70		1.12 ± 0.02	0.52 ± 0.03
90		0.53 ± 0.04	0.21 ± 0.02
100		0.39 ± 0.01	0.03 ± 0.001

*The data shows mean TP values ± standard deviation (n = 6)
0 % P reduced refers to unaltered media (TB and M9/YE)*

3.4.3 Effect of P minimisation in the fermentation media on *E. coli* biomass

3.4.3.1 Absorbance measurements

The growth of the recombinant *E. coli* strain was monitored by measuring the change in turbidity of the liquid cultures (section 3.3.2). The density of a cell suspension is

expressed as absorbance or optical density at 600 nm (OD_{600}), since for a limited range OD_{600} is directly proportional to the concentration of suspended cells [Cappucino and Sherman 2005]. However, at high absorbance values this relationship becomes imprecise due to secondary scattering [Lawrence and Maier 1977]. Table 3.5 shows the OD_{600} results of all liquid cultures after growth and recombinant protein expression of the model strain in the various media. The unaltered complex and semi-defined media yield similar OD_{600} values of 1.9 and 2.0 at the end of β -galactosidase expression, however, reduction of phosphorus in the fermentation media led to a modest decrease of final absorbancies. Statistical analysis (section 3.3.8) obtained a p value of < 0.001 which demonstrates that there is a statistically significant difference between the mean OD values of the different media containing the various amounts of P. This indicates that lowering the TP concentrations in fermentation media results in diminished cell growth of the model strain.

Table 3.5 Effect of P reduction in the fermentation media for the model *E. coli* strain upon optical density values (OD_{600}) observed after recombinant protein expression (6 h post induction)

% P reduction	OD values at 600 nm		
	LB	TB	M9/YE
0	1.9 ± 0.02	2.0 ± 0.03	1.9 ± 0.12
50		1.9 ± 0.02	1.9 ± 0.04
70		1.9 ± 0.03	1.7 ± 0.03
90		1.8 ± 0.05	1.2 ± 0.03
100		1.7 ± 0.04	0.8 ± 0.02

The data shows mean OD_{600} values ± standard deviation (n = 6); p < 0.001

0 % P reduced refers to unaltered media (TB and M9/YE)

3.4.3.2 Growth profiles of model *E. coli* strain for all fermentation media

The *E. coli* model strain was grown in all fermentation media employed during this study to determine if the phosphorus concentration in the media affects the growth kinetics of the bacteria. Growth curves of the model strain in LB, TB 0-100 % P reduced and M9/YE 0-100 % P reduced were obtained as outlined in section 3.3.4

and the results are illustrated in Figure 3.1 (complex media) and 3.2 (semi-defined media).

All six complex media, i.e. LB and TB 0-100 % P reduced, showed very similar growth profiles for the *E. coli* model strain. Cell growth started to level off after approximately 12 h in each of the media. During the exponential growth phase hours 4-8, the model strain showed consistently higher absorbance at 600 nm values in the TB media 50-100 % P reduced, however, after 14 h of growth difference in absorbance is less pronounced.

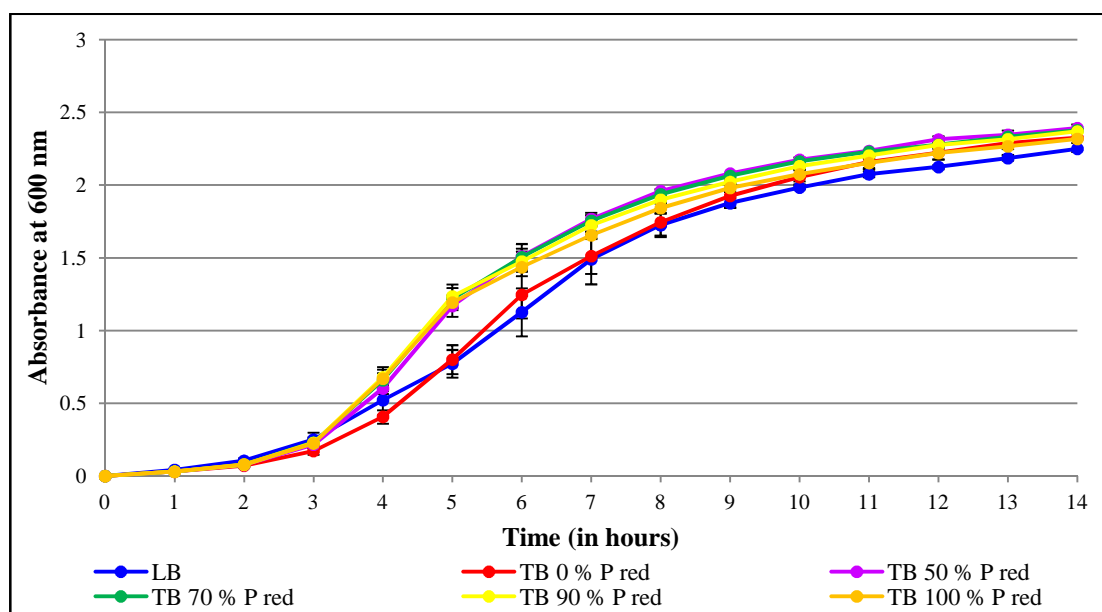


Figure 3.1 Growth curves of the model *E. coli* strain in complex fermentation media LB and TB 0-100 % P reduced

The data shows mean \pm standard deviation (n = 3)

LB: Luria-Bertani broth; TB: Terrific broth

Figure 3.2 shows the growth curves of the model strain in the M9/YE 0-100 % P reduced. From these findings it is clearly visible that reducing the phosphorus concentration in semi-defined media affects the absolute cell numbers produced as evident by the lower final absorbance at 600 nm values. However, P concentration in the media has no significant effect upon the kinetics of the model strain's growth, with stationary/near stationary levels achieved by approximately 7 h. For the first 4 h after inoculation the strain shows similar growth in all media but after 5 h cell division in M9/YE 100 % P reduced slows down and growth started to level off after 6-7 h resulting with the lowest final absorbance value for this media. The higher the

TP concentration in the semi-defined media the later the deceleration of cell division started and consequently the higher the final cell mass that was achieved as estimated from their absorbance values.

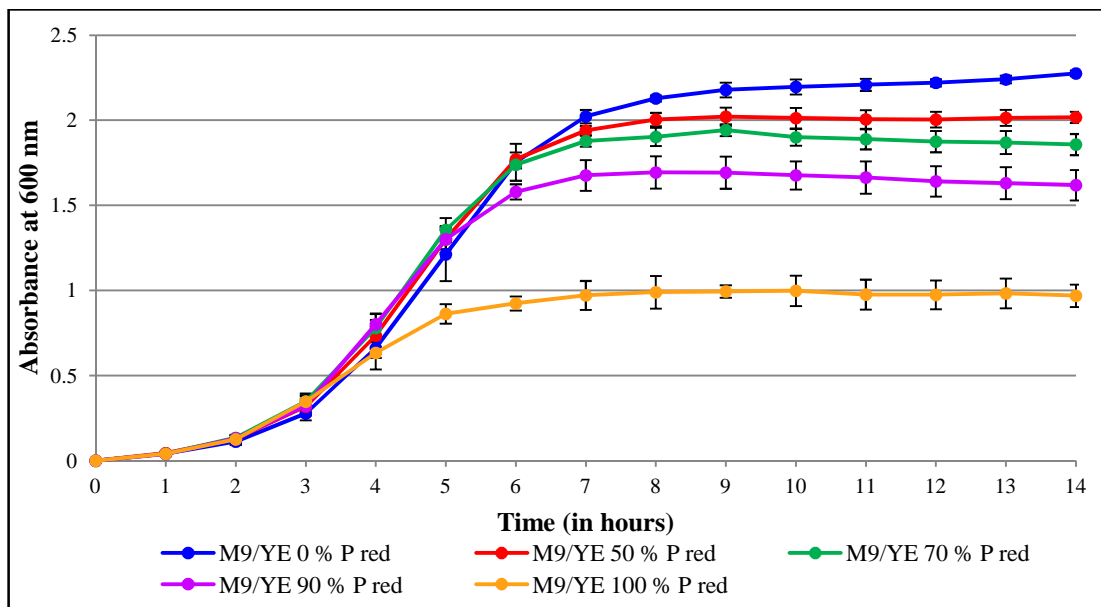


Figure 3.2 Growth curves of the model *E. coli* strain in semi-defined fermentation media M9/YE 0-100 % P reduced

The data shows mean \pm standard deviation ($n = 3$)

M9/YE: minimal media 9 supplemented with yeast extract

3.4.3.3 Dry cell weight

The effects of P reduction in the fermentation media for a model *E. coli* strain producing the recombinant protein on the overall biomass were evaluated in more detail through the determination of the dry cell weight (DCW). The photometric determination of bacterial concentration depends on light scattering rather than light absorption. The relationship between cell concentration and OD_{600} is only proportional for a limited range as secondary scattering occurs at increasing particle concentration [Lawrence and Maier 1977].

The DCW of the cultures expressed in the various media was determined as outlined in section 3.3.3. Table 3.6 represents the findings of stepwise P reduction upon cellular growth in the various media.

Table 3.6 Reduction of P and its effects on mean dry cell weight (DCW) values obtained after fermentation using the various media

% P reduction	Dry cell weight in g(DCW) l ⁻¹ (% mean in comparison to control)		
	LB	TB ^a	M9/YE ^b
0	0.76 ± 0.08	1.00 ± 0.09 (100.0)	0.87 ± 0.14 (100.0)
50		1.00 ± 0.06 (99.9)	0.83 ± 0.06 (94.9)
70		0.98 ± 0.09 (97.4)	0.83 ± 0.07 (95.7)
90		0.91 ± 0.09 (90.7)	0.57 ± 0.09 (65.6)
100		0.86 ± 0.10 (85.9)	0.27 ± 0.09 (30.7)

The data shows mean g(DCW) l⁻¹ values ± standard deviation (n = 12)

^a p < 0.001 (statistical analysis for all TB media); ^b p < 0.001 (statistical analysis for all M9/YE media)

Control for TB: 0 % P reduced TB media (unaltered media); control for M9/YE: 0 % P reduced M9/YE media (unaltered media)

The analysis showed that final cell concentration depends on the media used throughout fermentation. During this study the highest biomass of the model strain was obtained when employing unaltered TB with a mean DCW of 1.00 g(DCW) l⁻¹. This may be attributed to the increased concentration of nutrients, i.e. tryptone and yeast extract, present in this media formulation [Tripathi *et al.* 2009b]. M9/YE 0 % P reduced and LB produced cell densities of 0.87 and 0.76 g(DCW) l⁻¹, respectively. It was also determined that the reduction of P in TB as well as M9/YE affects the overall biomass that can be achieved during recombinant protein expression. Statistical analysis established that there is a significant difference (p < 0.001) between the mean gram of DCW per litre (g(DCW) l⁻¹) values for all the different TB and M9/YE media. However, the reduction of 100 % P from TB still yields more than 85 % of the mean g(DCW) l⁻¹ compared to the unaltered media. In addition, considering TB 0-70 % P reduced only, there is no statistical significant difference between the mean DCW values obtained (p = 0.591). P minimisation in the semi-defined media has a more considerable effect and reducing the P concentration by 90 % lead to a drop in biomass of 35 % and 100 % P reduced M9/YE only yielded about 30 % of the overall DCW. Nonetheless, a reduction of P up to 70 % does not adversely affect the DCW as there is no statistically significant difference between the mean DCW results of M9/YE 0 %, 50 % and 70 % P reduced (p = 0.498).

These findings show that it is possible to reduce the P concentration in TB and M9/YE up to 70 % without affecting the overall biomass yield but any further reduction can lead to a significant decrease of cellular mass.

3.4.3.4 pH analysis of all media

The reduction of overall biomass yields upon decrease of P in semi-defined media is most likely due to a drop in pH during growth of the cells and expression of the recombinant protein. Table 3.7 shows the pH values of the various media before inoculation with the model strain and after completion of recombinant protein expression.

Table 3.7 Mean pH values of all media employed during the study pre and post fermentation

	Media	pH before inoculation	pH after expression
COMPLEX MEDIA	LB	7.01 ± 0.02	6.89 ± 0.16
	TB 0 % P reduced	7.35 ± 0.03	7.22 ± 0.09
	TB 50 % P reduced	7.28 ± 0.04	7.15 ± 0.10
	TB 70 % P reduced	7.29 ± 0.03	6.98 ± 0.14
	TB 90 % P reduced	7.20 ± 0.02	7.04 ± 0.09
	TB 100 % P reduced	7.18 ± 0.04	6.97 ± 0.16
SEMI-DEFINED MEDIA	M9/YE 0 % P reduced	7.39 ± 0.02	6.81 ± 0.05
	M9/YE 50 % P reduced	7.40 ± 0.05	6.44 ± 0.05
	M9/YE 70 % P reduced	7.39 ± 0.04	6.18 ± 0.17
	M9/YE 90 % P reduced	7.32 ± 0.03	5.06 ± 0.04
	M9/YE 100 % P reduced	7.33 ± 0.05	4.69 ± 0.02

The data shows mean pH values ± standard deviation (n = 6)

0 % P reduced refers to unaltered media (TB and M9/YE)

All fermentation media were adjusted to pH 7.0-7.4, depending on the media recipe, prior to inoculation with the model strain, as the optimum pH range for *E. coli* growth is between pH 6.0-8.0 [Sambrook and Russell 2001; Schaechter 2010]. The pH of LB and all TB media pre and post fermentation remained relatively constant with pH values from 6.8 to 7.3. In the case of the semi-defined media at P reductions of 90 % and higher, the post fermentation media pH dropped to pH ≤ 5.06. *E. coli* produces acetic acid during aerobic fermentation and without the P salts as a

buffering component the pH in the media drops [Cappucino and Sherman 2005]. Ammonia containing chemicals such as ammonium chloride and ammonium sulphate also act as buffering compounds in media and their consumption during growth slowly decreases the buffering capacity of the media, thereby, resulting in a further pH drop [Krause *et al.* 2010]. At the point of inoculation the pH of the media was within the optimum pH range for the model strain but during the growth and expression of recombinant protein the pH dropped. Exposure of *E. coli* to low pH ranges has shown to induce the synthesis of acid shock proteins, which most likely are responsible for the induction of the bacteria's acid tolerance response (ATR) [Šeputienė *et al.* 2006]. Therefore, there are some viable cells present at the end of expression but at much lower concentrations compared to the unaltered semi-defined media.

The complex ingredients in LB and TB contain amino acids, peptones, and proteins, which are amphoteric compounds and thereby act as natural buffers [Cappucino and Sherman 2005] preventing a considerable change in pH.

3.4.4 Effect of P minimisation in the fermentation media on the recombinant protein

3.4.4.1 Concentration of the recombinant protein

The model strain expresses β -galactosidase in two fractions, as soluble functional protein and as non-soluble misfolded inclusion bodies. Electrophoretic analysis of *E. coli* MC1061 carrying the pProEx-HTb plasmids without the *lacZ*-HT insert showed no visible bands with molecular masses of approximately 116 kDa. Therefore, the concentrations (mg g(DCW)^{-1}) for both types of the recombinant protein as well as the total protein concentration of the combined fractions could be determined using SDS-PAGE analysis in combination with densitometry (section 3.3.7).

A standard calibration curve relating β -galactosidase concentration (purified enzyme from *E. coli*) to area density on an SDS-PAGE was produced and a typical gel is illustrated in Figure 3.3. All of the standard calibration curve gels and the subsequent gels run during the P minimisation study contained two standard β -galactosidase

concentrations (0.75 and 1.5 μg loaded per lane). These standards were used as controls to confirm successful electrophoresis and staining of the gels as well as ensure correlation to the standard curve.

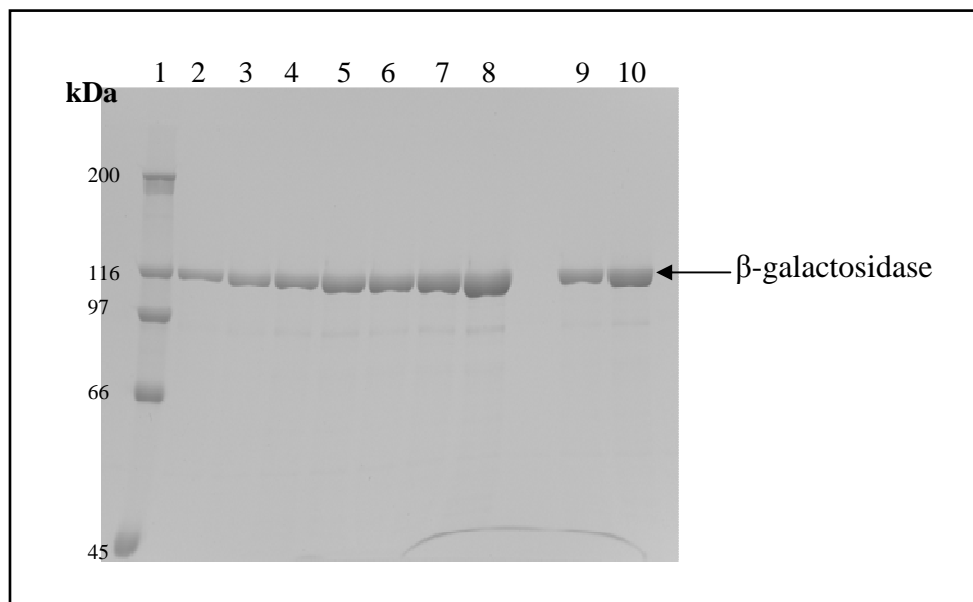


Figure 3.3 SDS-PAGE analysis of standard β -galactosidase concentration for the production of the calibration curve

Lane 1: Molecular mass marker (Sigma-Aldrich); Lanes 2-8: commercial β -galactosidase loaded at increasing protein amounts (0.5 μg per lane – 2.0 μg per lane); Lanes 9 & 10: commercial β -galactosidase at standard concentrations of 0.75 μg and 1.5 μg protein loaded per lane.

The standard calibration curve of β -galactosidase concentration versus area density on an SDS-PAGE gel is shown in Figure 3.4. The r^2 value of 0.997 indicates a good linear relationship between protein concentration loaded per lane and area density of the resulting band agreeing with previous literature reported [Wheater and Cook 2000].

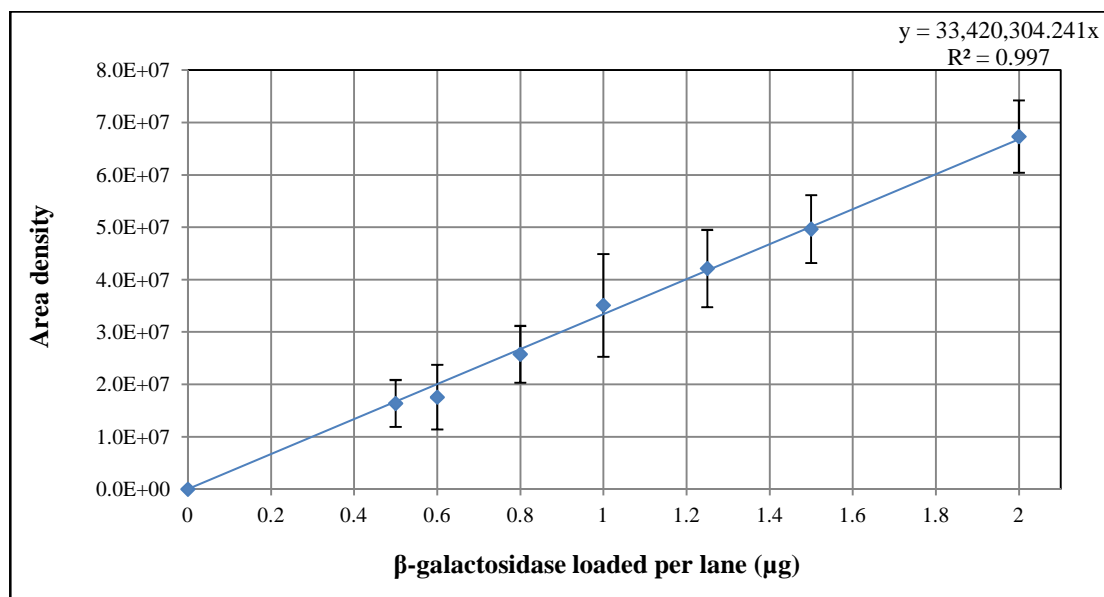


Figure 3.4 Standard calibration curve relating β -galactosidase loaded per lane on an SDS-PAGE gel and the area density of the resulting band

The data shows mean area density value \pm the standard deviation ($n = 6$)

The soluble, non-soluble and total protein fractions, obtained at the end of fermentation for each of the media, were analysed and the recombinant β -galactosidase concentrations could be established using the area density of the bands in correlation with the standard curve.

Table 3.8 shows the results of β -galactosidase concentrations obtained for the soluble fractions post bacterial growth and protein expression. The data indicates that the highest concentration of soluble functional recombinant protein is produced in LB with a mean β -galactosidase concentration of $12.4 \text{ mg}(\beta\text{-gal}) \text{ g}(\text{DCW})^{-1}$. This is followed by the unaltered TB media and then the 0 % P reduced semi-defined media with mean concentration values of 11.2 and $10.2 \text{ mg}(\beta\text{-gal}) \text{ g}(\text{DCW})^{-1}$, respectively. The results also show that there is no statistically significant difference between concentrations of soluble β -galactosidase per $\text{g}(\text{DCW})$ obtained when using the TB media containing varying amounts of P ($p = 0.810$). Additionally, a p value of 0.098 suggests there is no statistically significant difference between the mean values of soluble β -galactosidase concentrations obtained for the two types of complex media, i.e. the same amount of soluble functional enzyme is produced by the model strain in LB and TB.

A reduction of P up to 70 % in M9/YE does not give rise to a statistically significant difference in the mean concentration values for soluble β -galactosidase per g(DCW) ($p = 0.508$). However, any further reduction led to a drop in soluble recombinant protein concentration. This is also most likely due to the decrease in media pH during fermentation. A severe change in pH because of the accumulation of acetate impairs cell growth and can have a negative effect on recombinant product formation [Krause *et al.* 2010; Hortsch and Weuster-Botz 2011]. Reduction in media pH leads to significant instability of plasmids, such as pProEx-HTb carrying the *lacZ* gene in the model *E. coli* strain, which results in plasmid loss and thereby a decrease in protein yield. In addition, low pH in fermentation media can also cause degradation of ampicillin causing a loss of these ampicillin-resistant plasmids in the bacterial cells allowing further increase in cell growth without protein production or a very low protein yield as observed for the 100 % P reduced M9/YE [Sivashanmugam *et al.* 2009].

Table 3.8 Reduction of P in the fermentation media and its effects on soluble recombinant protein concentrations

% P reduced	Mean concentration of soluble β -galactosidase (mg(β -gal) g(DCW) ⁻¹)		
	(% mean in comparison to control)		
	LB	TB ^a	M9/YE ^b
0	12.4 \pm 1.5	11.2 \pm 1.6 (100.0)	10.2 \pm 0.9 (100.0)
50		10.2 \pm 0.8 (90.8)	10.9 \pm 1.3 (106.4)
70		10.3 \pm 2.1 (92.0)	10.6 \pm 0.5 (103.6)
90		10.4 \pm 1.6 (92.8)	7.7 \pm 1.5 (75.4)
100		10.7 \pm 1.7 (95.2)	5.8 \pm 1.2 (56.9)

The data shows mean value \pm the standard deviation with $n = 6$

^a $p = 0.810$ (statistical analysis for all TB media); ^b $p < 0.001$ (statistical analysis for all M9/YE media)

Control for TB: 0 % P reduced TB media (unaltered media); control for M9/YE: 0 % P reduced M9/YE media (unaltered media)

Figure 3.5 illustrates typical SDS-PAGE gels obtained during the analysis of the soluble protein fractions. The reduction of P in the fermentation media did not appear to have any visible effects on the protein band patterns obtained for the soluble fractions of the model strain expressed in complex or semi-defined media with varying amounts of TP.

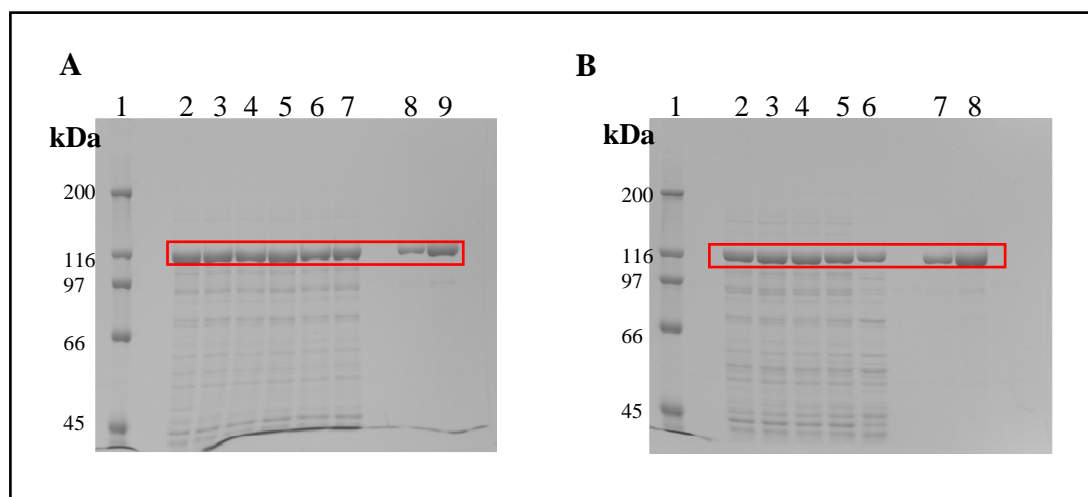


Figure 3.5 SDS-PAGE analysis of the soluble fractions for determination of the recombinant protein concentrations

A

Lane 1: Molecular mass marker (Sigma-Aldrich); Lane 2: soluble fraction of model strain expressed in LB; Lane 3: soluble fraction of model strain expressed in unaltered TB; Lane 4: soluble fraction of model strain expressed in TB 50 % P reduced; Lane 5: soluble fraction of model strain expressed in TB 70 % P reduced; Lane 6: soluble fraction of model strain expressed in TB 90 % P reduced; Lane 7: soluble fraction of model strain expressed in TB 100 % P reduced; Lane 8: 0.75 μg β -galactosidase standard; Lane 9: 1.5 μg β -galactosidase standard. β -Galactosidase is highlighted in red.

B

Lane 1: Molecular mass marker (Sigma-Aldrich); Lane 2: soluble fraction of model strain expressed in unaltered M9/YE; Lane 3: soluble fraction of model strain expressed in M9/YE 50 % P reduced; Lane 4: soluble fraction of model strain expressed in M9/YE 70 % P reduced; Lane 5: soluble fraction of model strain expressed in M9/YE 90 % P reduced; Lane 6: soluble fraction of model strain expressed in M9/YE 100 % P reduced; Lane 7: 0.75 μg β -galactosidase standard; Lane 8: 1.5 μg β -galactosidase standard. β -Galactosidase is highlighted in red.

The results for the β -galactosidase concentrations in the non-soluble fractions of the model strain grown and expressed in all media are illustrated in Table 3.9. The highest concentration of non-soluble protein produced by the model strain was obtained during fermentation in unaltered TB ($6.4 \text{ mg}(\beta\text{-gal}) \text{ g}(\text{DCW})^{-1}$). Employing LB or M9/YE results in similar inclusion body formation with a mean β -galactosidase concentration of $3.5 \text{ mg}(\beta\text{-gal}) \text{ g}(\text{DCW})^{-1}$ for both media. The recombinant protein production conditions used during this study achieve high levels of β -galactosidase expression in all selected fermentation media. It is likely, that the high local concentrations of the product with insufficient amount of folding-promoting proteins yields the formation of inclusion bodies [Sørensen and Mortensen 2005]. TB was specifically devised to achieve high levels of protein

expression [Tartoff and Hobbs 1987], however, these high production rates of a non-native protein have led to the increased accumulation of inclusion bodies compared to the other media.

A 50 % reduction of P salts added to both the defined and complex media does not produce a statistically significant difference in the mean concentrations of non-soluble β -galactosidase per g(DCW) obtained ($p = 0.072$ and $p = 0.291$, respectively). Any further P minimisation has a significant effect on the concentration of non-soluble β -galactosidase resulting in a notable decrease of non-soluble protein per g(DCW). The reduction of P in the media may result in a marginally reduced overall protein synthesis rate which slightly decreases the overall protein concentration but allows the same amount of soluble, functional protein to be formed.

Table 3.9 Reduction of P in the fermentation media and its effects on non-soluble recombinant protein concentrations

% P reduced	Mean concentration of non-soluble β -galactosidase (mg(β -gal) g(DCW) ⁻¹) (% mean in comparison to control)		
	LB	TB ^a	M9/YE ^b
0	3.5 \pm 0.8	6.4 \pm 1.4 (100.0)	3.5 \pm 0.2 (100.0)
50		5.1 \pm 0.7 (80.0)	3.7 \pm 0.5 (106.7)
70		4.3 \pm 1.4 (67.0)	2.9 \pm 0.5 (84.5)
90		2.5 \pm 0.7 (39.4)	0.9 \pm 0.2 (25.4)
100		2.7 \pm 1.1 (42.1)	0.4 \pm 0.2 (10.3)

The data shows the mean value \pm the standard deviation ($n = 6$)

^a $p < 0.001$ (statistical analysis for all TB media); ^b $p < 0.001$ (statistical analysis for all M9/YE media)

Control for TB: 0 % P reduced TB media (unaltered media); control for M9/YE: 0 % P reduced M9/YE media (unaltered media)

Typical gels obtained during the determination of the non-soluble β -galactosidase concentration are illustrated in Figure 3.6. The gels for the complex as well as the semi-defined media clearly show that there is a decrease in non-soluble protein concentration per g(DCW) upon reduction of P in the fermentation media. However, there are no visible changes in the protein band patterns.

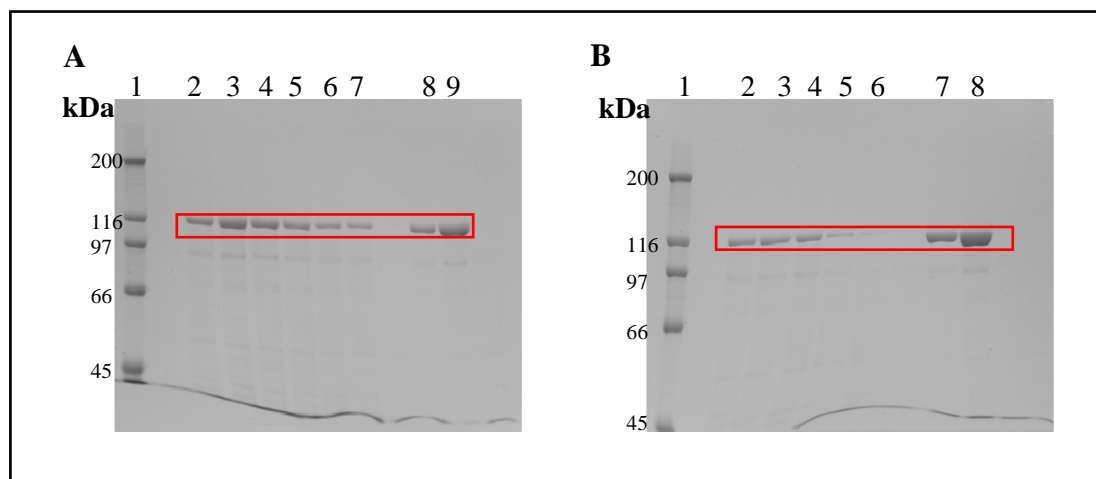


Figure 3.6 SDS-PAGE gels of the non-soluble fractions of the model strain after fermentation during the P minimisation study

A

Lane 1: Molecular mass marker (Sigma-Aldrich); Lane 2: non-soluble fraction of model strain expressed in LB; Lane 3: non-soluble fraction of model strain expressed in unaltered TB; Lane 4: non-soluble fraction of model strain expressed in TB 50 % P reduced; Lane 5: non-soluble fraction of model strain expressed in TB 70 % P reduced; Lane 6: non-soluble fraction of model strain expressed in TB 90 % P reduced; Lane 7: non-soluble fraction of model strain expressed in TB 100 % P reduced; Lane 8: 0.75 μg β -galactosidase standard; Lane 9: 1.5 μg β -galactosidase standard. β -Galactosidase is highlighted in red.

B

Lane 1: Molecular mass marker (Sigma-Aldrich); Lane 2: non-soluble fraction of model strain expressed in unaltered M9/YE; Lane 3: non-soluble fraction of model strain expressed in M9/YE 50 % P reduced; Lane 4: non-soluble fraction of model strain expressed in M9/YE 70 % P reduced; Lane 5: non-soluble fraction of model strain expressed in M9/YE 90 % P reduced; Lane 6: non-soluble fraction of model strain expressed in M9/YE 100 % P reduced; Lane 7: 0.75 μg β -galactosidase standard; Lane 8: 1.5 μg β -galactosidase standard. β -Galactosidase is highlighted in red.

The results for the total β -galactosidase concentrations obtained in complex and semi-defined media are shown in Table 3.10. TB media achieved higher total recombinant protein concentrations with a mean value of $18.2 \text{ mg}(\beta\text{-gal}) \text{ g}(\text{DCW})^{-1}$ compared to LB with $17.0 \text{ mg}(\beta\text{-gal}) \text{ g}(\text{DCW})^{-1}$ and $13.2 \text{ mg}(\beta\text{-gal}) \text{ g}(\text{DCW})^{-1}$ for unaltered M9/YE media. This is due to the increased concentration of non-soluble recombinant enzyme found in TB (Table 3.9). The findings also show a slight decrease in mean total recombinant protein upon reduction of exogenous P added to TB which is caused by the reduced amount of non-soluble protein only. However, this reduction is not statistically significant ($p = 0.370$). In M9/YE the reduction of TP results in a notable decrease in the concentration of recombinant protein produced. However, a statistically significant difference occurs only if the exogenous

P salts added are reduced by more than 70 %. A P reduction of ≤ 70 % yields a p value of 0.508 indicating there is no statistical significant difference between the mean concentrations of the total β -galactosidase per g(DCW). The results for the soluble and non-soluble β -galactosidase fractions (Tables 3.8 and 3.9) both indicated that a reduction of TP in the semi-defined media of 90 and 100 % leads to a decrease in protein concentration. Therefore, the reduced amount of β -galactosidase in the total protein fractions is due to both a reduction in soluble and non-soluble recombinant protein.

Table 3.10 Reduction of P in the fermentation media and its effects on total recombinant protein concentrations

% P reduced	Mean concentration of total β -galactosidase (mg(β -gal) g(DCW) ⁻¹) (% mean in comparison to control)		
	LB	TB ^a	M9/YE ^b
0	17.0 \pm 2.9	18.2 \pm 2.4 (100.0)	13.2 \pm 0.5 (100.0)
50		17.5 \pm 2.1 (96.5)	13.4 \pm 0.7 (101.7)
70		17.5 \pm 2.8 (96.1)	12.8 \pm 1.2 (97.4)
90		15.9 \pm 2.8 (87.6)	8.5 \pm 1.5 (64.3)
100		15.8 \pm 2.2 (86.7)	5.5 \pm 0.6 (41.6)

The data shows the mean value \pm the standard deviation (n = 6)

^a p = 0.370 (statistical analysis for all TB media); ^b p < 0.001 (statistical analysis for all M9/YE media)

Control for TB: 0 % P reduced TB media (unaltered media); control for M9/YE: 0 % P reduced M9/YE media (unaltered media)

Figure 3.7 shows typical gels obtained during the analysis of the total protein fraction from the model strain expressed in complex media and semi-defined media containing varying amounts of P. There are also no visible changes in the protein band patterns obtain for the various media.

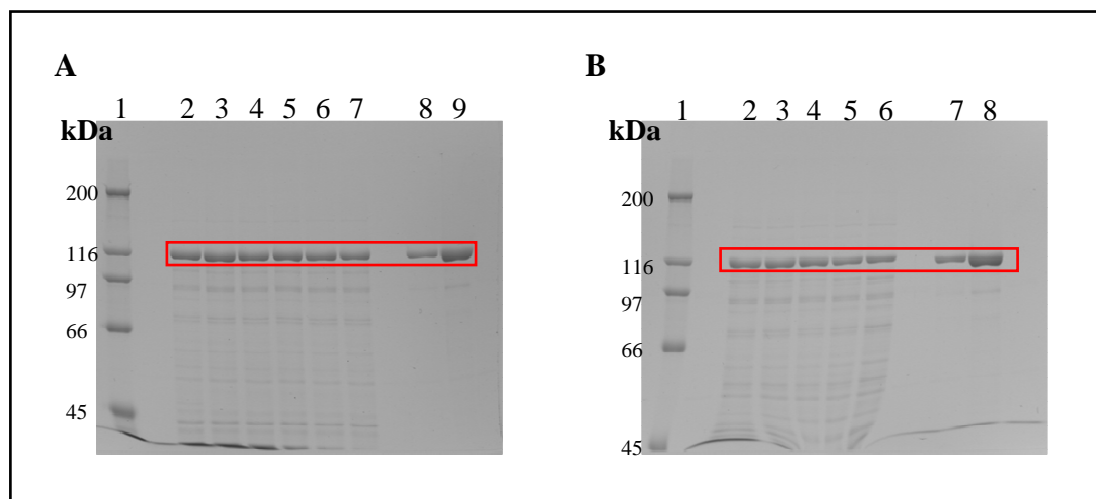


Figure 3.7 Electrophoretic analysis of the total protein fractions from the model strain obtained during the P minimisation study

A

Lane 1: Molecular mass marker (Sigma-Aldrich); Lane 2: total protein fraction of model strain expressed in LB; Lane 3: total protein fraction of model strain expressed in unaltered TB; Lane 4: total protein fraction of model strain expressed in TB 50 % P reduced; Lane 5: total protein fraction of model strain expressed in TB 70 % P reduced; Lane 6: total protein fraction of model strain expressed in TB 90 % P reduced; Lane 7: total protein fraction of model strain expressed in TB 100 % P reduced; Lane 8: 0.75 μg β -galactosidase standard; Lane 9: 1.5 μg β -galactosidase standard. β -Galactosidase is highlighted in red.

B

Lane 1: Molecular mass marker (Sigma-Aldrich); Lane 2: total protein fraction of model strain expressed in unaltered M9/YE; Lane 3: total protein fraction of model strain expressed in M9/YE 50 % P reduced; Lane 4: total protein fraction of model strain expressed in M9/YE 70 % P reduced; Lane 5: total protein fraction of model strain expressed in M9/YE 90 % P reduced; Lane 6: total protein fraction of model strain expressed in M9/YE 100 % P reduced; Lane 7: 0.75 μg β -galactosidase standard; Lane 8: 1.5 μg β -galactosidase standard. β -Galactosidase is highlighted in red.

3.4.4.2 Functional activity of recombinant protein

It was established from the DCW determination and the recombinant protein quantification that a reduction of up to 70 % of P salts in both TB and M9/YE was possible without adversely affecting biomass or protein product yields. The routine industrial use of such P-reduced media would only be applied in practice if the decrease in TP did not negatively affect the recombinant protein's functional quality. The effect of P minimisation in the fermentation media on the quality of the recombinant protein produced was assessed by determining the enzyme units present in the soluble crude extracts as well as by establishing the kinetic parameters K_m and V_{max} for the crude recombinant β -galactosidase (section 3.3.5). The results of the

mean enzyme activity values are expressed as enzyme units per milligram of soluble β -galactosidase (EU mg(soluble β -gal)⁻¹) in Table 3.11. There is no statistically significant difference between the mean β -galactosidase activities from any of the crude extracts of the model strain expressed in complex media ($p = 0.440$). The results also show that there is no statistically significant difference between the mean enzyme units obtained for M9/YE with the varying concentrations of P ($p = 0.127$). However, there is a notable difference between the mean values obtained for the complex media compared to the mean values for the semi-defined media. The findings indicate that higher enzyme activity per gram of soluble recombinant β -galactosidase could be observed in the crude extracts of the cells grown in complex media. An immediate explanation is not obvious and kinetic analysis was carried out on the soluble crude extracts to further evaluate the effects of P minimisation in fermentation media.

Table 3.11 β -Galactosidase activity values of the soluble crude fraction obtained during the P minimisation study

% P reduced	Mean β -galactosidase activity (EU mg(soluble β -gal) ⁻¹)		
	(% mean in comparison to control)		
	LB	TB ^a	M9/YE ^a
0	2180 \pm 50	2128 \pm 111 (100.0)	1694 \pm 85 (100.0)
50		2242 \pm 135 (105.3)	1804 \pm 189 (106.5)
70		2169 \pm 63 (101.9)	1794 \pm 94 (105.9)
90		2185 \pm 70 (102.7)	1834 \pm 355 (108.3)
100		2188 \pm 83 (102.8)	1996 \pm 53 (117.8)

The data shows the mean value \pm the standard deviation ($n = 6$)

^a $p = 0.440$ (statistical analysis for all TB media); ^b $p = 0.127$ (statistical analysis for all M9/YE media)

Control for TB: 0 % P reduced TB media (unaltered media); control for M9/YE: 0 % P reduced M9/YE media (unaltered media)

Comparative kinetic analysis was used for the additional assessment of the recombinant β -galactosidase quality. The kinetic parameters (K_m and V_{max}) were determined from the soluble crude extract of the model strain expressed in LB, the unaltered media TB and M9/YE media and both media with 70 % P reduced. The results of this analysis are illustrated in Table 3.12.

Table 3.12 Effect of P reduction on recombinant protein's kinetic parameters K_m and V_{max}

% P reduction	LB		TB		M9/YE	
	K_m^a (mM)	V_{max}^b (mM min ⁻¹)	K_m^a (mM)	V_{max}^b (mM min ⁻¹)	K_m^a (mM)	V_{max}^b (mM min ⁻¹)
0	0.69 ± 0.09	26.81 ± 3.44	0.76 ± 0.11	28.21 ± 3.07	0.84 ± 0.05	27.66 ± 1.67
70			0.66 ± 0.14	24.77 ± 2.72	0.68 ± 0.01	26.35 ± 2.84

The data shows the mean value ± the standard deviation ($n = 4$)

^a $p = 0.080$ (statistical analysis for all K_m values); ^b $p = 0.501$ (statistical analysis for all V_{max} values)

The findings indicate that the K_m and V_{max} values recorded are not influenced by type of media or the amount of P salts present in it ($p = 0.080$ for K_m and $p = 0.501$ for V_{max}). The K_m value for *E. coli* β -galactosidase with ONPG as substrate was previously determined to be 0.161 mM [Wallenfels and Malhotra 1960]. The slight difference in the constant may be due to the fact that crude enzyme solutions were used during this comparative study instead of purified β -galactosidase.

These results indicate that expression of β -galactosidase in P-reduced media does not negatively impact on the quality of the recombinant protein.

3.5 Conclusion

The findings obtained during the P minimisation study provide laboratory-scale proof of principle that a reduction of P salts added to fermentation media of up to 70 % is possible without negatively impacting on the biomass yield or recombinant protein produced. P is not a constituent of amino acids which are assembled to form proteins, nor is it subsequently integrated as a post-translational modification (PTM) in any commercial protein produced in *E. coli*. Therefore, it is likely the results presented would be generally applicable to the manufacturing of most/all recombinant proteins in *E. coli*. P is an essential core cell nutrient, consequently, TP concentration in the media will almost certainly affect primary cellular biomass production. In addition, P salts play a role as buffering components in the fermentation media and a reduction in media pH has been shown to decrease the recombinant protein production independently of cell biomass formation. Previous studies have shown that, for *E. coli*, the generalised relationship between growth rate and nutrient consumption is similar between shake-flask experiments and batch fermentation [Li *et al.* 1990; Luli and Strohl 1990; Aristidou *et al.* 1999]. During scale-up experiments with the 70 % P-reduced fermentation media variations of process parameters such as modifications in growth conditions (e.g. exact nutrient composition, temperature, dissolved oxygen concentration) or fermentation format/scale will be investigated and while these will almost certainly affect the biomass yield attained, the effect upon the ratio of protein to biomass produced will be minimal once the media pH is kept stable.

A 70 % reduction in P salts added to fermentation media can potentially bear significant practical applications during large-scale commercial recombinant protein production. To place the results into an industrial context, 70 % P reductions translate into P savings of 12.5 to 16.2 kg per 10,000 l batch depending on the type of media employed. Such reductions can give rise to significant annual P savings for the entire biopharmaceuticals sector. Additionally, this will result in a much cleaner manufacturing process, placing less pressure on the waste stream treatment facilities while also reducing the ultimate risk of eutrophication. Furthermore, the use of P-

reduced media achieves savings in manufacturing cost and can assist in slowing the rate of depletion of natural P sources.

Chapter Four - A comparative environmental analysis of traditional multi-chromatographic versus affinity-based purification for a recombinant protein and reduction of the environmental impact of the affinity-based purification scheme

4.1 Introduction

Waste minimisation has become an important aspect in the manufacturing of therapeutic proteins in order to minimise the possible impact that the process may have on the environment as well as the costs associated with the process. The preferred means of waste management is the prevention or reduction of wastes (e.g. phosphorus) [New *et al.* 2000] and chapter three detailed the potential savings in phosphorus wastes that can be achieved when employing P-reduced fermentation media.

This chapter investigates the environmental impacts of downstream processing as well as potential waste minimisation strategies that could result in a reduction or elimination of P in the waste streams. Downstream processing, which entails protein purification, is the second stage of biopharmaceutical production and traditionally at industrial-scale it consists of three to four chromatography techniques in combination with membrane filtration steps [Carta and Jungbauer 2010]. On average 100-1,000 l of buffer per kg of product are required for each of those unit operations resulting in large waste stream volumes [Ho 2010]. Many of the buffers employed during protein isolation contain inorganic and organic salts including compounds with N and P [GE Healthcare 2010b; Junker 2010]. As part of this chapter the potential environmental impact savings when using recombinant proteins with polyhistidine tags in combination with an immobilised metal affinity chromatography (IMAC)-based purification scheme were compared to traditional multi-chromatographic protein recovery. In addition, P waste minimisation studies were undertaken by reducing and replacing the P salts present in purification buffers.

4.2 Materials

As described in section 2.2. In addition, all chromatographic media were purchased from GE Healthcare, Uppsala, Sweden.

4.3 Experimental methods

4.3.1 Fermentation and cell lysis

Fermentation of the model *E. coli* strain and expression of the recombinant protein was carried out in 500 ml conical flasks with culture volumes of 100 ml using the optimised expression conditions (section 2.4.8) and 70 % P-reduced TB media (section 3.4.2). Upon completion of expression, the cells were harvested by centrifugation at 7,500 rpm, 4 °C and the cell pellets were stored at -80 °C.

4.3.2 Traditional multi-chromatographic purification scheme

The test-tube methods for determination of starting pH were carried out according to the method described by GE Healthcare [2013] and Williams and Frasca [2001]. Aliquots of crude enzyme were mixed with small quantities of ion exchange resin (DEAE Sepharose CL-6B), previously equilibrated to the various pH values and the supernatant was assayed for β -galactosidase activity. The following buffers were used during the test: Tris-maleate/sodium hydroxide (pH 6.0-7.0) and Tris-Cl (pH 7.5-8.0).

The model strain cell pellets were resuspended in ion exchange chromatography (IEC) buffer A (50 mM Tris-maleate/NaOH, 50 mg l⁻¹ trypsin inhibitor, pH 6.0), lysed using sonication (30 s bursts for 20 min with intermittent cooling) and subsequently centrifuged (12,000 rpm, 4 °C, 20 min). The resulting protein solution was purified using the BioLogic LP System with model 2110 fraction collector (BioRad Laboratories, Hercules, U.S.) at 4 °C. The crude protein mixture was applied to a DEAE Sepharose CL-6B column (column volume (CV) 8.0 ml), pre-equilibrated with IEC A. Unbound protein was removed by washing with 5 CV of IEC A at a flow rate of 1.0 ml min⁻¹. The remaining proteins were eluted in 20 CV IEC A with a 0-0.8 M NaCl gradient at 1.0 ml min⁻¹ and collected in 3.0 ml fractions. Any remaining bound protein was subsequently eluted by washing with IEC B

(50 mM Tris-maleate/NaOH, 2.0 M NaCl, 50 mg l⁻¹ trypsin inhibitor, pH 6.0). All fractions were assayed for enzyme activity according to section 2.3.9. The protein concentration was determined by absorbance at 280 nm and the absorbance units were recorded by the Biologic LP dataview software. Fractions with β -galactosidase activity were pooled, concentrated (typically by 11-14 fold) and diafiltered (UF/DF) into TEV buffer (50 mM Tris-Cl, 0.5 mM EDTA, 1 mM DTT, pH 8.0) using a 50 ml Amicon Stirred Cell ultrafiltration unit (Millipore, Cork) with Millipore 100,000 Da nitrocellulose membrane. AcTev protease (Life Technologies) was added to the concentrate according to manufacturing specification to remove the His₆-tag and the solution was incubated at 4 °C for four days. Ammonium sulphate (3.0 M) was added to the cleaved protein sample to a final concentration of 0.2 M and the solution loaded onto 6 ml Phenyl Sepharose HP pre-equilibrated with hydrophobic interaction chromatography (HIC) buffer A (50 mM sodium phosphate, 0.2 M ammonium sulphate, 50 mg l⁻¹ trypsin inhibitor, pH 7.0). Unbound protein was removed via washing with 5 CV of HIC A at a flow rate of 1.0 ml min⁻¹ and the β -galactosidase enzyme was eluted with 10 CV of HIC B (50 mM sodium phosphate, 50 mg l⁻¹ trypsin inhibitor, pH 7.0) using 1.0 ml min⁻¹. The partially purified protein was concentrated by ultrafiltration (UF) and then loaded onto a Superdex200 column (CV 125 ml), pre-equilibrated with size exclusion chromatography (SEC) A buffer (50 mM sodium phosphate, 150 mM NaCl, 1 mM DTT, 50 mg l⁻¹ trypsin inhibitor, pH 6.5). Protein was eluted with 1.2 CV SEC A buffer at 0.3 ml min⁻¹ and the effluent was collected in 2 ml fractions. Fractions were tested for enzyme activity and protein concentration as described previously and fractions containing β -galactosidase activity were pooled. Total β -galactosidase activity was determined as outlined in section 2.3.9 and total protein concentration was established using the Bradford assay (section 3.3.5).

4.3.3 Affinity-based purification scheme

The model strain cell pellets were resuspended in immobilised metal affinity chromatography (IMAC) A buffer (50 mM sodium phosphate, 500 mM NaCl, 40 mM imidazole, 50 mg l⁻¹ trypsin inhibitor, pH 7.5), lysed using sonication (30 s bursts for 20 min with intermittent cooling) and subsequently centrifuged at 12,000 rpm, 4 °C for 20 min. Small-scale batch studies were carried out to determine

the optimum metal ion employed during IMAC purification when using Chelating Sepharose Fast Flow and the recombinant β -galactosidase. Aliquotes of the crude enzyme were mixed with 1.0 ml of charged resin previously pre-equilibrated with IMAC A and the supernatant was tested for β -galactosidase activity. The following metal ions were tested: nickel (Ni^{2+}), copper (Cu^{2+}), cobalt (Co^{2+}) and zinc (Zn^{2+}).

The crude protein sample was applied to a Ni^{2+} -charged Chelating Sepharose Fast Flow column (CV 5 ml), pre-equilibrated with IMAC A. Unbound protein was removed by washing with 12 CV of IMAC A using a flow rate of 1.0 ml min^{-1} . The protein of interest was eluted using 15 CV of IMAC B (50 mM sodium phosphate, 500 mM NaCl, 500 mM imidazole, 50 mg l^{-1} trypsin inhibitor, pH 7.5) at a flow rate of 1.0 ml min^{-1} and the effluent was collected in 3.0 ml fractions. The fractions were tested for enzyme activity and protein concentration as outlined previously (section 4.3.2). Fractions with β -galactosidase activity were pooled, concentrated (typically by 7-8.5 fold) and diafiltered (UF/DF) into TEV buffer. AcTev protease was added to remove the His₆-tag and the solution was incubated at 4 °C for four days. The cleaved protein solution was loaded onto Superdex200 column (CV 125 ml), pre-equilibrated with SEC A buffer. Protein was eluted from the SEC column with 1.2 CV SEC A buffer at a flow rate of 0.3 ml min^{-1} and the effluent collected in 2.0 ml fractions. These fractions were tested for β -galactosidase activity and protein concentration and the fractions with the protein of interest were pooled. The total β -galactosidase activity was established according to section 2.3.9 and total protein concentration was determined as described in section 3.3.5.

4.3.4 Chemical oxygen demand

Chemical oxygen demand (COD) was determined using the closed reflux, colorimetric method (5220 D) [Eaton and Franson 1995]. A sample (2.5 ml) was placed in a 10 ml culture tube to which 1.5 ml digestion solution and 3.5 ml sulphuric acid reagent were added. The digestion solution was prepared by adding 10.216 g potassium dichromate ($\text{K}_2\text{Cr}_2\text{O}_7$), previously dried at 103 °C for 2 h, 167 ml concentrated sulphuric acid (H_2SO_4), and 33.3 g mercury (II) sulphate (HgSO_4) to 500 ml of distilled water. Once all chemicals were dissolved and the solution cooled to room temperature, it was diluted to 1 l with distilled water. Sulphuric acid reagent

contained silver sulphate (Ag_2SO_4) which was added to concentrated H_2SO_4 at a rate of 5.5 g of Ag_2SO_4 per kg of H_2SO_4 . The culture tubes were sealed, placed in a block digester (Hach Company, Düsseldorf, Germany) and refluxed for 2 h at 150 °C. Once cooled to room temperature, the sample absorbance was read at 600 nm with a UV-visible spectrophotometer (Hach Company, Düsseldorf, Germany).

A standard curve relating oxygen concentration to absorbance at 600 nm was constructed to quantify oxygen consumed in the reaction. COD standards, ranging from 20-900 $\text{mg}(\text{O}_2) \text{ l}^{-1}$, were prepared in triplicate by diluting a stock solution of potassium hydrogen phthalate 10,000 $\text{mg}(\text{O}_2) \text{ l}^{-1}$ in distilled water (425 mg potassium hydrogen phthalate ($\text{HOOC}_6\text{H}_4\text{COOK}$) dissolved in 500 ml of distilled water). The absorbance of the standard solutions were also determined according to the closed reflux, colorimetric method (5220 D) [Eaton and Franson 1995].

4.3.5 Ammonia

The ammonia content of the buffers was determined by using the preliminary distillation step (4500-NH₃ B) and the titrimetric method (4500-NH₃ C) as outlined in the standard method for examination of water and wastewater [Eaton and Franson 1995]. The analysis entailed the distillation of 9.2 ml of sample with 185 ml water and 25 ml borate buffer solution. The borate buffer contained 88 ml of 0.1 M sodium hydroxide (NaOH) which was added to 0.025 M sodium tetraborate ($\text{Na}_2\text{B}_4\text{O}_7$) solution and diluted to 1 l with distilled water. The pH of the sample mixture was adjusted to pH 9.5 using NaOH (6M). This solution was distilled using a Büchi distillation unit B-324 (BÜCHI Labortechnik AG, Flawil, Switzerland) and ammonia was collected in 18.5 ml indicating boric acid solution. This boric acid solution contained 20 g boric acid (H_3BO_3) and 10 ml mixed indicator solution (200 mg methyl red indicator in 100 ml 95 % isopropanol (IPA) was added to 100 mg methyl blue in 50 ml IPA) which were mixed and diluted to 1 l with distilled water). During the final step the ammonia was titrated against H_2SO_4 (0.01 M) until the indicator turned a pale lavender colour. The volume of titrant used is employed in a formula as described in Eaton and Franson [1995] to determine the concentration of ammonia in the sample.

4.3.6 Organic nitrogen

The theoretical organic nitrogen (N_{org}) was calculated using the chemical structures of the salts contained in each buffer which were prepared as outlined in Dawson *et al.* [1986].

4.3.7 Minimisation of total phosphorus (TP) in waste streams produced during recombinant protein purification

A waste minimisation strategy was put in place for the affinity-based purification system. The recombinant protein was purified according to section 4.3.3 but the SEC load was divided into four aliquots and each sample was eluted from the SEC column with sodium phosphate salt concentration in the SEC A buffer ranging from 5 to 50 mM. The resulting elution fractions were analysed for β -galactosidase activity according to section 2.3.9 and subsequently converted to a percentage of the total amount of β -galactosidase loaded during each SEC run.

In addition, alternative buffer salts were investigated to replace the sodium phosphate in IMAC and SEC. The following salts were analysed for TP, COD and theoretical N according to sections 3.3.7, 4.3.5 and 4.3.6, respectively, at a concentration of 50 mM: ACES, ADA, BES, Bis-Tris, Bis-Tris propane, HEPES, MES, PIPES, Clark and Lubs buffer (KH_2PO_4 and NaOH) [Dawson *et al.* 1986], Tricine and Tris-Acetate. The optimal buffering salt selected (50 mM ACES) was used as chromatographic solvent in the affinity-based purification scheme to isolate the recombinant β -galactosidase. The recombinant protein was purified as outlined in section 4.3.3 with the following buffers: IMAC C (50 mM ACES, 500 mM NaCl, 40 mM imidazole, 50 mg l⁻¹ trypsin inhibitor, pH 7.5), IMAC D (50 mM ACES, 500 mM NaCl, 500 mM imidazole, 50 mg l⁻¹ trypsin inhibitor, pH 7.5), TEV and SEC B (50 mM ACES, 150 mM NaCl, 1 mM DTT, 50 mg l⁻¹ trypsin inhibitor, pH 6.5). The total β -galactosidase activity and total protein concentration at each step were established according to section 2.3.9 and 3.3.5, respectively.

4.4 Results and Discussion

4.4.1 Purification of recombinant β -galactosidase using a traditional multi-chromatographic scheme

A relevant literature review was carried out to identify an optimal multi-chromatographic purification approach for recombinant β -galactosidase. The traditional purification scheme selected, consisting of ion exchange chromatography (IEC) followed by hydrophobic interaction chromatography (HIC) and completed with size exclusion chromatography (SEC) as outlined in section 4.3.2, is the recommended strategy for purifying untagged proteins [GE Healthcare 2010b]. This purification strategy is a model system broadly applicable for many proteins and not just for the recombinant β -galactosidase protein.

The initial chromatographic step (IEC) separates the proteins in the model strain crude extract based on their differences in overall charge. To evaluate the optimal starting condition for IEC the theoretical isoelectric point (pI) of the recombinant protein was determined bioinformatically according to section 2.3.2 and found to be 5.34. The starting buffer pH for anion exchangers should be approximately 1 pH unit above the pI of the protein [GE Healthcare 2013]. The test-tube method, performed as described in section 4.3.2, identified the optimum pH for binding to the DEAE Sepharose CL-6B to be pH 6.0 which is close to 1 pH unit above the theoretical pI of recombinant β -galactosidase.

Figure 4.1 is a typical chromatogram for the initial purification step of the recombinant protein using a buffer pH of 6.0 and DEAE Sepharose CL-6B resin. The crude protein solution was loaded and the β -galactosidase absorbed onto the column. Contaminating proteins were removed during a washing step and the protein of interest was subsequently eluted by an ascending salt gradient from 0 to 0.8 M NaCl over a total volume of 160 ml. The β -galactosidase protein started to elute from the column at a NaCl concentration of 0.2 M as a sharp peak between fractions 30 and 46 which were combined for further purification. It was evident from the chromatogram (Figure 4.1) that some recombinant β -galactosidase was eluted during the wash which contributed to the decrease in percent yield observed during this step (yield after IEC step 30.8 %; Table 4.1). Reducing the amount of crude protein

solution loaded onto the IEC column did not change the percentage of recombinant β -galactosidase lost during the wash step.

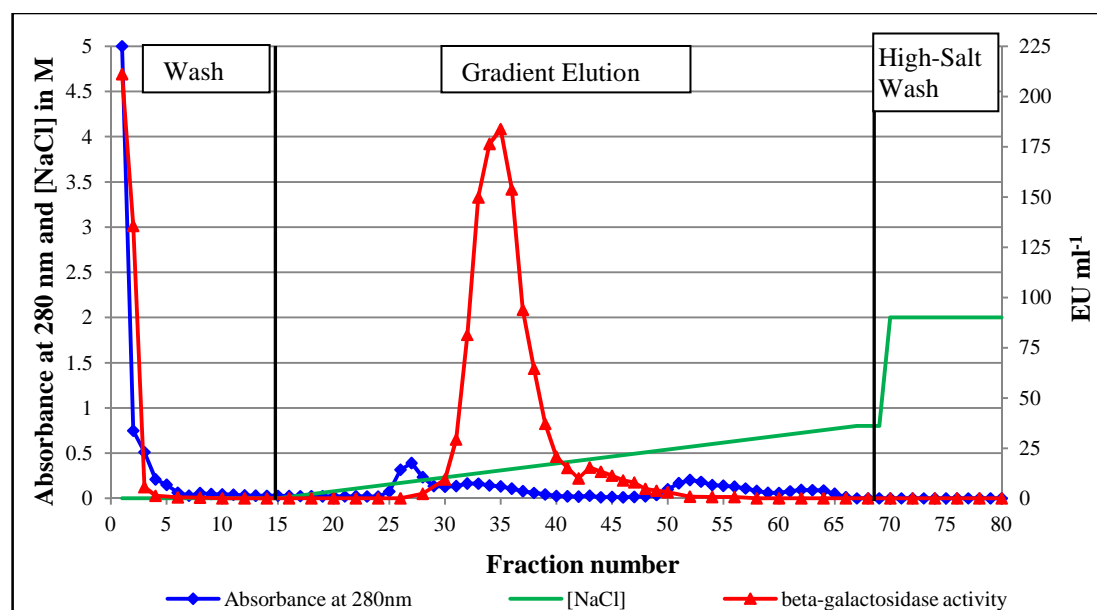


Figure 4.1 Chromatogram for the purification of recombinant β -galactosidase using an IEC column containing DEAE Sepharose CL-6B

The pooled IEC fractions with enzyme activity were concentrated and the buffer was exchanged prior to His₆-tag removal and HIC using ultrafiltration in combination with diafiltration (UF/DF). This method was selected as it is widely employed in biotechnology and biopharmaceutical manufacturing for the concentration of proteins and buffer exchange. UF/DF is gentle on enzyme activity and is associated with high recovery rates and relatively fast processing times [Phillips and Signs 2001; Soetaert and Vandamme 2010]. It was carried out with a 100 kDa membrane and a 12.4 fold concentration was achieved. The concentrated protein solution was incubated with TEV protease for polyhistidine tag removal. By employing the plasmid pProEx-HTb, the recombinant protein is expressed with an N-terminal His₆-tag as well as a TEV cleavage recognition site allowing removal of the tag via a TEV protease [EMBL 2012]. Cleavage was monitored via SDS-PAGE and Western blot analysis to ensure complete removal of the His₆-tag. Table 4.1 shows an increase in yield after the UF/DF and tag cleavage step. UF/DF of the sample removed any excess salts and possibly other protein inhibitors that may have interfered with the enzyme activity. β -Galactosidase activity can be inhibited by heavy metals, organomercuric compounds and various alcohols and amines. The enzymatic activity also depends on the presence of divalent ions (in particular Mg²⁺ and Ca²⁺),

therefore, metal chelators binding these ions can affect the total enzyme activity detected [Wallenfels and Weil 1972]. The removal of the His₆-tag may have helped with overall protein stability or the availability of the divalent ions, thereby, further increasing the enzyme activity.

The partially purified and digested protein was applied to a Phenyl Sepharose HP column for hydrophobic interaction chromatography. Trial HIC runs showed that the protein of interest bound to the chromatographic media at ammonium salt concentrations of ≥ 0.2 M which were therefore selected as the starting conditions for HIC. A typical chromatogram for HIC during the traditional multi-chromatography purification of the recombinant β -galactosidase is shown in Figure 4.2. The protein mixture was absorbed onto the column and no β -galactosidase activity was detected in the wash fractions while contaminating proteins were removed (Figure 4.2 and Table 4.1). The bound proteins were eluted from the column via one step using an ammonium salt-free buffer resulting in one clear protein peak which coincides with the β -galactosidase activity peak.

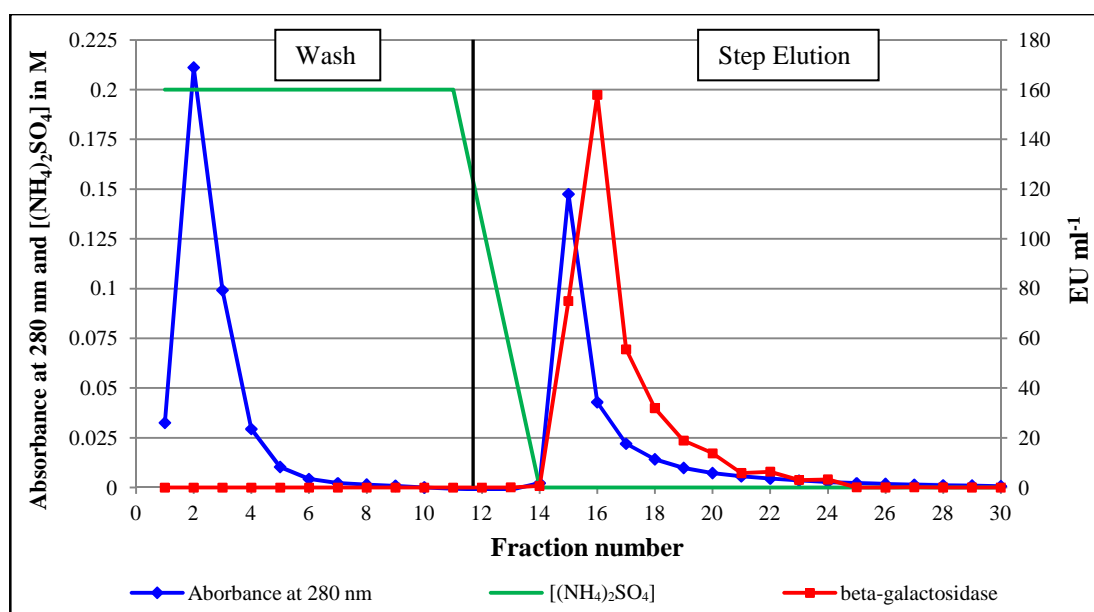


Figure 4.2 Chromatogram for the purification of recombinant β -galactosidase by HIC using Phenyl Sepharose HP

The pooled elution fractions from HIC were concentrated 15 fold using ultrafiltration (UF) with a 100 kDa membrane. The final step in the traditional multi-chromatographic purification scheme was size exclusion chromatography using

Superdex200 resin. The fractionation range of this medium is 10 to 600 kDa (globular proteins) which includes the molecular mass of the recombinant protein (464 kDa). A typical chromatogram for SEC is illustrated in Figure 4.3. There are four protein peaks evident in the chromatogram and the major peak (fractions 29-34) overlaps with the enzyme activity peak observed. The first protein peak (fractions 25-28) also displays β -galactosidase activity and corresponds to higher order oligomeric forms of the recombinant protein judged from native PAGE analysis. These oligomers are commonly observed during recombinant β -galactosidase expression [Juers *et al.* 2000]. Sufficient resolution of the two peaks was achieved by applying a flow rate of 0.3 ml min^{-1} and loading a sample volume of 1.2 % of the total column volume. Any further reduction in flow rate led to peak tailing of the main β -galactosidase peak and an increased concentration of the sample prior to SEC resulted in interference of the separation due to viscosity. Fractions 29-34 were combined as they contained the purified protein of interest, recombinant β -galactosidase in its tetrameric, functional form.

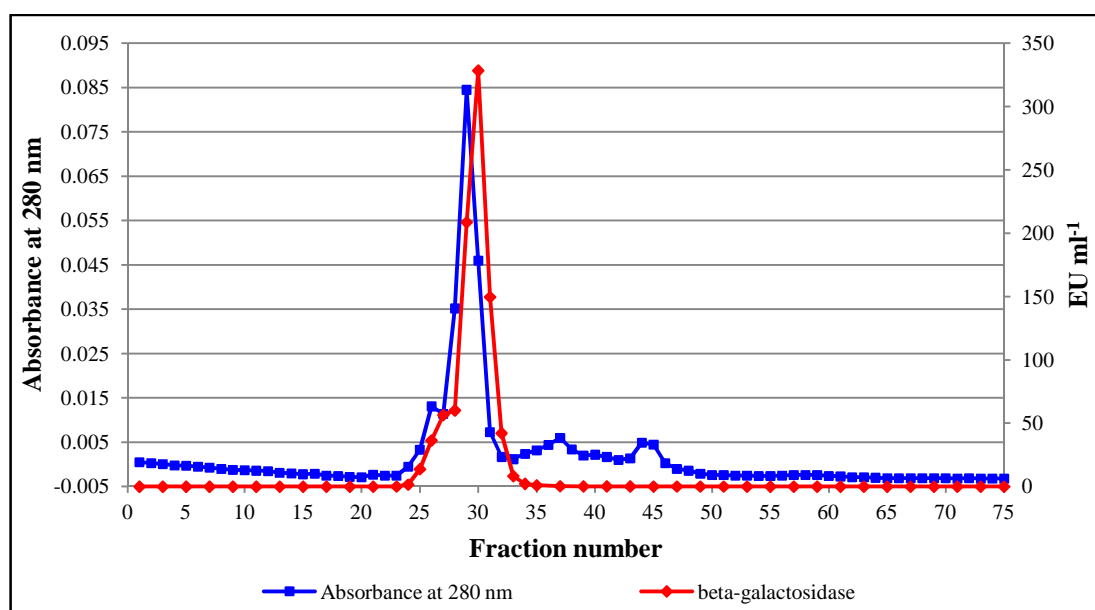


Figure 4.3 Chromatogram for the final step of the traditional multi-chromatographic scheme for recombinant β -galactosidase purification by SEC using Superdex200

Table 4.1 Purification table for recombinant β -galactosidase purified using the traditional multi-chromatographic scheme

Step	Total volume (ml)	Total protein (mg)	Total activity (EU)	Specific activity (EU mg ⁻¹)	Yield (%)	Purification factor
Crude extract	3.0	26.8	13,627.6	508.5	100	1
IEC peak	51.0	8.60	4,198.0	488.1	30.8	0.96
UF/DF	4.1	7.23	5,961.7	824.6	43.7	1.62
TEV-cleaved protein	4.1	9.84	6,459.6	656.5	47.4	1.29
HIC peak	24	2.64	2,821.2	1,068.6	20.7	2.10
UF	1.6	2.37	2,530.2	1,067.6	18.6	2.10
SEC peak	12	1.09	1,535.5	1,408.7	11.3	2.77

The results of the traditional multi-chromatographic purification scheme are shown in Table 4.1. Using the outlined combination of methods resulted in a purification factor of 2.77 and a final yield of 11.3 % for the recombinant β -galactosidase enzyme. The yield and purification factor fall within the wide ranges typically recorded when proteins are purified by such means [Chakraborti *et al.* 2000; McKinney *et al.* 2004; Zhang *et al.* 2004]. The purification factor necessary to achieve protein isolation to homogeneity is relatively low which is due to the high level expression of the recombinant protein in the model *E. coli* strain.

Crude and purified protein samples of recombinant β -galactosidase were analysed by SDS- and native PAGE as outlined in section 2.3.8. The purified protein migrated as a single band during SDS-PAGE analysis, indicating purification to homogeneity. The protein band had an apparent molecular mass of 116.6 kDa estimated with respect to the protein standards of known molecular masses (Figure 4.4 A). This corresponds to the theoretical molecular mass of the recombinant β -galactosidase monomer after the N-terminal His₆-tag was removed as determined according to section 2.3.2. A single band could be observed for recombinant β -galactosidase during native PAGE analysis when stained for protein and activity (Figure 4.4 B). This shows that the enzyme was purified to homogeneity and the higher order oligomers were removed during the SEC step. Moreover, the enzyme activity was confirmed by zymogram and native PAGE overlay technique using the purified protein (Figure 4.4 B).

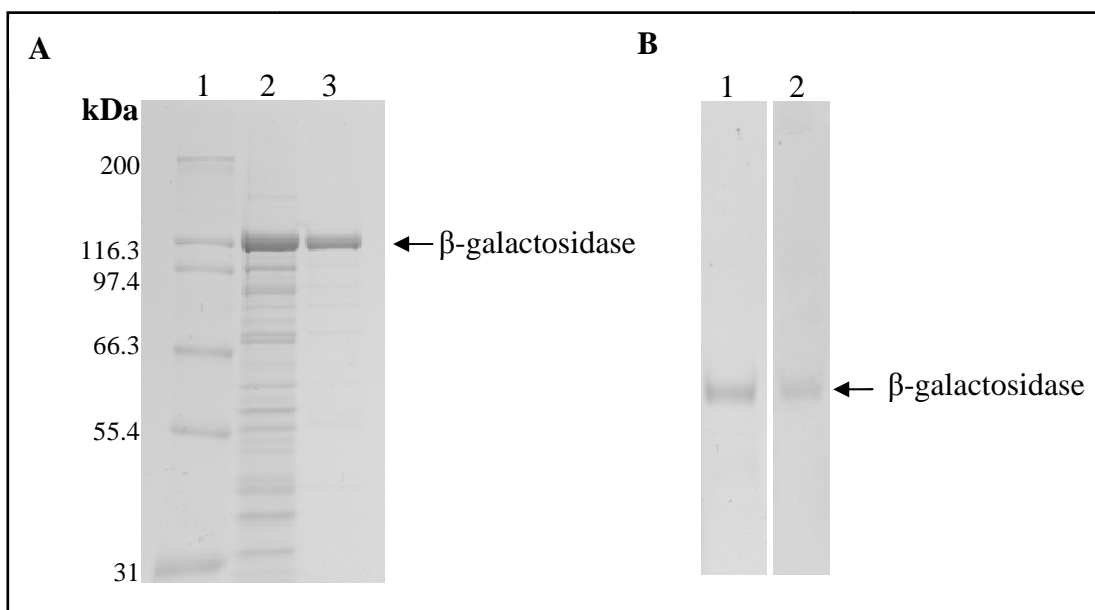


Figure 4.4 Electrophoretic analysis of recombinant β -galactosidase purified using traditional multi-chromatographic scheme

A

SDS-PAGE analysis stained for protein – Lane 1: Protein Standard Mark 12 (Life Technologies); Lane 2: crude protein preparation; Lane 3: purified protein.

B

Native PAGE analysis of purified enzyme stained for protein (Lane 1) and stained for activity (Lane 2).

In general, a protein purification strategy consists of three levels. The first step is the initial purification of the target molecule from the source material which in this case is carried out via IEC. The next stage (intermediate purification) involves the removal of bulk contaminants and was achieved using HIC during the purification of this recombinant product. The last level is a polishing step which focuses on final removal of trace contaminants and adjustments of pH, salts and additives for use or storage [GE Healthcare 2010b]. During this study the trace contaminants, including higher order oligomers of the β -galactosidase, were removed during the SEC step.

The traditional multi-chromatographic purification scheme was developed based on the general approach for untagged proteins [GE Healthcare 2010b] but also through establishing the level of purity after each optimised step. Thereby, this purification set-up was a six-step scheme. Pre-existing published approaches for *E. coli* β -galactosidase purification consists of an initial ammonium sulphate fractionation step after cell lysis, followed by IEC and subsequent concentration via ultrafiltration or

ammonium sulphate precipitation. The partially purified protein was then applied to an SEC column. To improve the final purity of the enzyme some methods added an affinity chromatography step prior to SEC while other protocols used an additional IEC after SEC [Fowler and Zabin 1983; Cupples *et al.* 1990; Jacobson and Matthews 1992; Juers *et al.* 2000]. This indicates that five steps are required to achieve homogeneity when purifying *E. coli* β -galactosidase via traditional methods. The additional step employed during this purification strategy is the TEV cleavage step to remove the His₆-tag. The TEV cleavage step was included to make it directly comparable to the affinity-based system and to ensure both purification strategies yielded an identical β -galactosidase protein.

4.4.2 Purification of recombinant β -galactosidase using an affinity-based scheme

The model strain used during this study expresses the recombinant β -galactosidase with an N-terminal His₆-tag which facilitates the purification by immobilised metal affinity chromatography (IMAC). At a pH \geq 7.0 the imidazole side chains of the histidine residues have a net negative charge and bind to the positively charged metal ions chelated onto the chromatography media [Petty 2001]. Due to the presence of higher order β -galactosidase oligomers and other minor contaminants as determined by SDS-PAGE analysis the recombinant protein was further purified by SEC. This affinity-based purification scheme for the recombinant His₆-tagged product is outlined in section 4.3.3.

A typical chromatogram for the purification of recombinant β -galactosidase from the crude protein solution of the model strain on Ni²⁺-charged IMAC is shown in Figure 4.5. In general, Ni²⁺ is the preferred metal for IMAC purification [GE Healthcare 2007] and the small-scale batch purification with the recombinant enzyme, described in section 4.3.3, showed the highest binding capacity to the chromatography resin charged with this metal ion.

The IMAC step resulted in a purification factor of 2.23 which together with the chromatogram (Figure 4.5) clearly indicates the removal of contaminating proteins during the wash step. While a small amount of recombinant β -galactosidase was also lost during washing, the majority (54.2 % recovery) was eluted and collected during

the step elution as seen by the protein with overlapping enzyme activity peak in Figure 4.5.

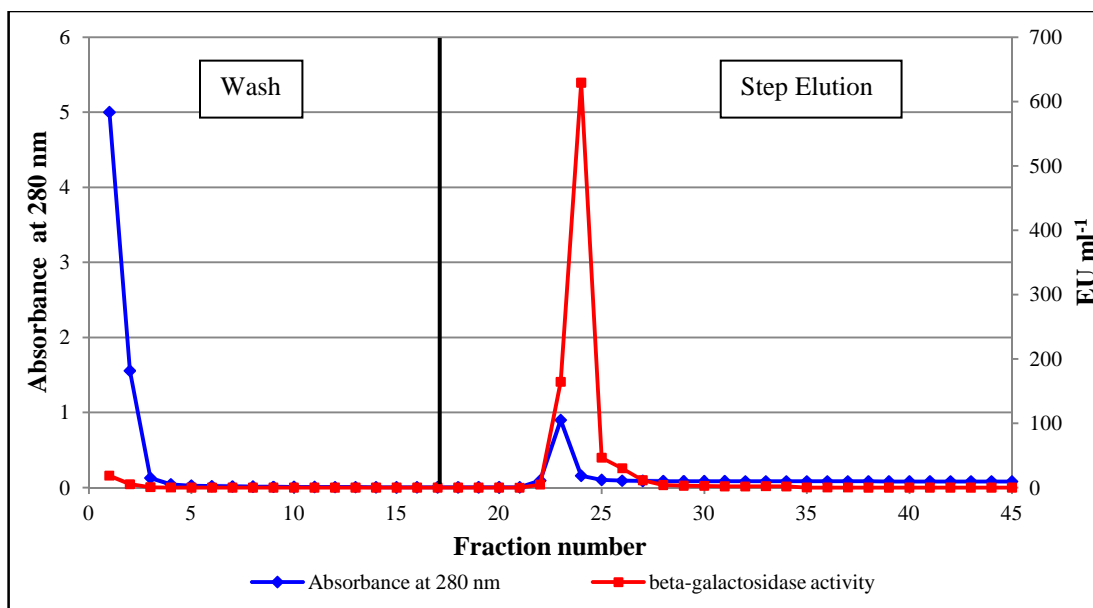


Figure 4.5 Chromatogram for the purification of recombinant β -galactosidase with IMAC containing Ni^{2+} -charged Chelating Sepharose Fast Flow

The partially purified protein was concentrated 7.8 fold and the buffer exchanged via UF/DF for subsequent His₆-tag cleavage. Cleavage was monitored using SDS-PAGE and Western blot analysis to ensure complete removal of the tag. The digested protein mixture was loaded onto the SEC column to remove any remaining contaminants as well as higher order oligomers of the protein of interest. Figure 4.6 shows a typical chromatogram of an SEC run during the purification of recombinant β -galactosidase. There are six protein peaks evident in the chromatogram two of which overlap with the two enzyme activity peaks. The first protein and β -galactosidase activity peak (fractions 23-28) corresponds to higher order oligomeric forms of the recombinant product as determined via native PAGE analysis. The second peak (fractions 29-35) coincides with the protein of interest, i.e. recombinant β -galactosidase in tetrameric, functional form.

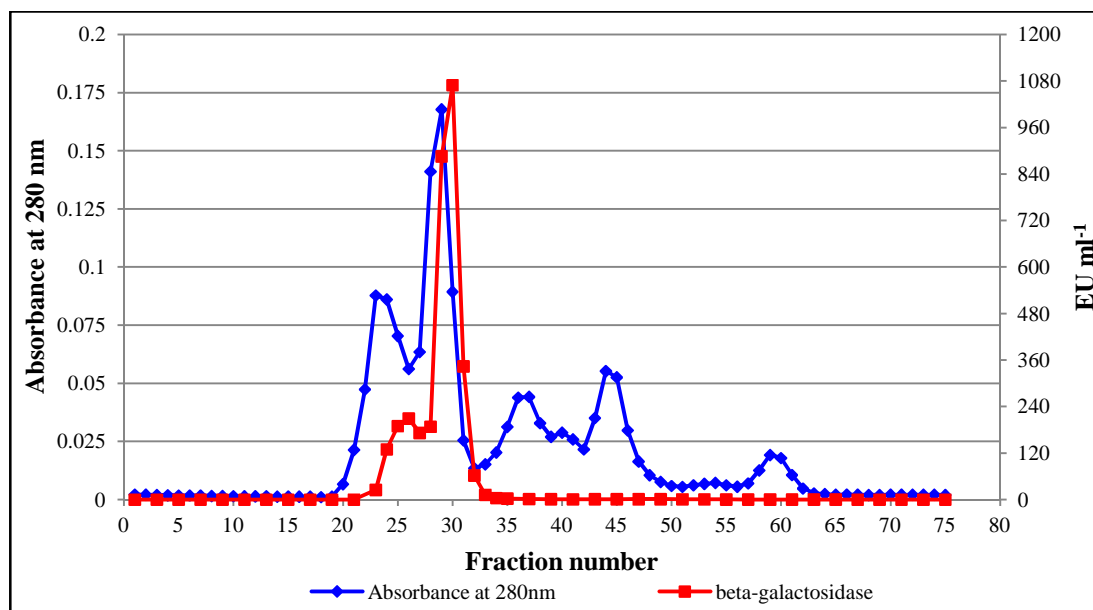


Figure 4.6 Chromatogram for the final step of the affinity-based scheme for recombinant β -galactosidase purification by SEC using Superdex200

The affinity-based scheme yielded a purification factor of 3.02 and a yield of 34.6 % for the recombinant β -galactosidase (Table 4.2). The yield and purification factor for polyhistidine-tagged recombinant proteins can vary greatly but the results obtained fall within the recorded range [Das *et al.* 2005; Maestro *et al.* 2008; Marx *et al.* 2008].

Table 4.2 Purification table for recombinant β -galactosidase purified using the affinity-based chromatography scheme

Step	Total volume (ml)	Total protein (mg)	Total activity (EU)	Specific activity (EU mg ⁻¹)	Yield (%)	Purification factor
Crude extract	2.75	23.21	12,365.6	532.8	100	1
IMAC peak	23.5	5.64	6,698.3	1,187.6	54.2	2.23
UF/DF	3.0	6.00	7,302.7	1,217.1	59.1	2.28
TEV-cleaved protein	3.0	6.60	9,149.7	1,386.3	74.0	2.60
SEC peak	14.0	2.66	4,273.0	1,606.4	34.6	3.02

Crude and purified protein samples isolated using the affinity-based scheme of recombinant β -galactosidase were analysed by SDS- and native PAGE as outlined in

section 2.3.8. The purified enzyme migrated as a single band during SDS-PAGE analysis (Figure 4.7 **A**), indicating purification to homogeneity. The protein band had an apparent molecular mass of 116.6 kDa, estimated with respect to the protein standards of known molecular masses. This corresponds to the theoretical molecular mass of the recombinant β -galactosidase monomer (without His₆-tag) as determined according to section 2.3.2. The native PAGE gels of the purified β -galactosidase stained for protein and enzyme activity (Figure 4.7 **B**) show a single band illustrating that the protein was purified to homogeneity and the higher order oligomers were removed during the SEC step. In addition, the zymogram and native PAGE overlay technique confirmed enzyme activity of the purified product (Figure 4.7 **B**).

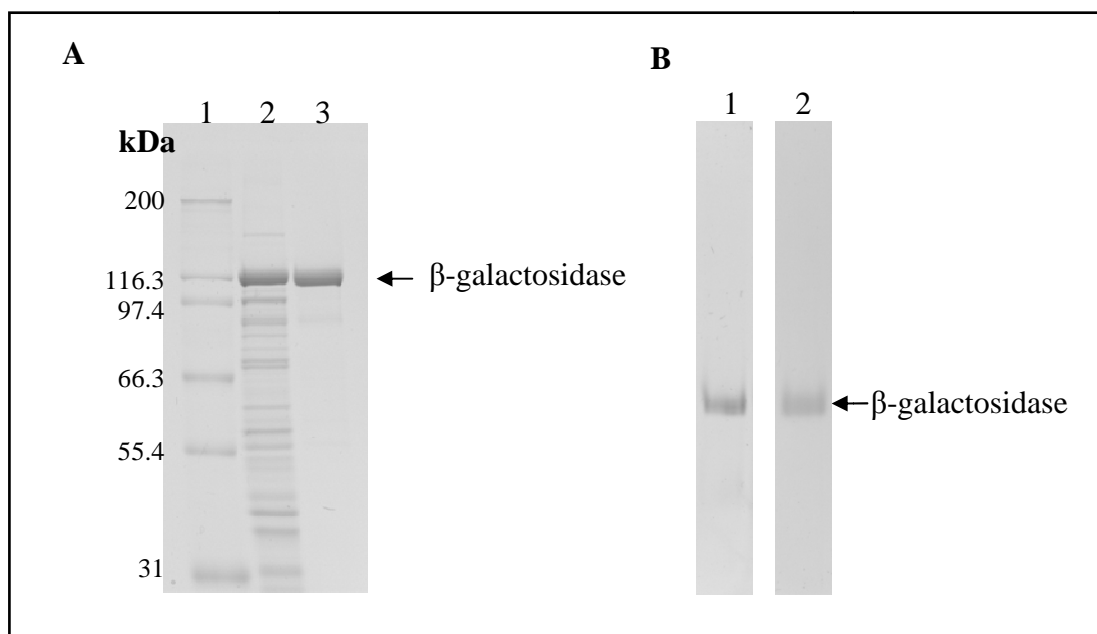


Figure 4.7 Electrophoretic analysis of recombinant β -galactosidase purified using the affinity-based chromatography scheme

A

SDS-PAGE analysis stained for protein – Lane 1: Protein Standard Mark 12 (Life Technologies); Lane 2: crude protein preparation; Lane 3: purified protein.

B

Native PAGE analysis of purified enzyme stained for protein (Lane 1) and stained for activity (Lane 2).

The use of IMAC as the first initial purification level allows the reduction of chromatographic methods and decreases the overall steps required from six to four while achieving the same purification. The polyhistidine tag facilitates a much higher specificity for the recombinant protein during IMAC compared to the anion

exchange media which binds all proteins in the mixture with similar charge characteristics. The separation resolution that can be achieved with IEC depends on the amount of proteins with the same adsorption characteristics (pI and charge distribution) as the target protein [Gräslund *et al.* 2000]. Link *et al.* [1997] have shown that more than 90 % of *E. coli* proteins have a pI between 4.0 and 7.0. As the recombinant β -galactosidase lies within that range (pI = 5.34) like many other proteins, it is expected that separation of the target protein from contaminants via IEC is limited and further steps are required. IMAC eliminates this problem due to the high affinity of the His₆-tag to the Ni²⁺-charged resin which allows the capture of the recombinant β -galactosidase and removal of the major contaminants in one step.

Reducing the number of purification steps resulted in an increase in yield from 11.3 to 34.6 % for the recombinant β -galactosidase product. Moreover, utilising IMAC not only aids protein recovery but decreasing the number of steps involved also reduces overall buffer consumption. The buffers employed for both purification processes are standard Tris- and phosphate-based buffers commonly used during the purification of proteins [GE Healthcare 2010b]. All these buffers contain low to significant levels of phosphorus, ammonia, organic nitrogen and COD levels which if released into the environment can considerably influence the water quality in the surrounding water bodies [Khan and Ansari 2005].

4.4.3 Environmental analysis of both purification schemes

An environmental analysis was carried out on all the buffers employed during the two purification strategies to evaluate the different environmental factors (TP, NH₃, N_{org} and COD) for each scheme. These particular parameters were selected as they are indicators of significant pollution and are routinely monitored in the wastes produced from manufacturing facilities [Carty *et al.* 1997; Hammer and Hammer 2004]. The analyses were carried out according to sections 3.3.7 (TP), 4.3.4 (COD), 4.3.5 (NH₃) and 4.3.6 (N_{org}) and the results are shown in Table 4.3.

The findings illustrate that, unsurprisingly, the highest concentration of TP could be detected in the buffers containing phosphate salts, i.e. HIC A, HIC B, SEC A, IMAC A and IMAC B. While the buffers employed for IEC (IEC A and IEC B) as well as the TEV buffer did not contain any P salts, low concentrations of TP were still

detected due to the presence of the trypsin inhibitor. This was added to all buffers to prevent proteolysis of the recombinant protein to a characteristic 90 kDa degradation product. The trypsin inhibitor used during this study is clarified from *Glycine max* (soybean) and can contain up to 10 % sodium phosphate buffer salts [Kunitz 1945; Sigma-Aldrich 2013b].

Table 4.3 Environmental analysis of all buffers employed during the purification of the recombinant β -galactosidase

Buffer	TP (mg(P) l ⁻¹)*	NH ₃ (mg(NH ₃) l ⁻¹)*	N _{org} (mg(N _{org}) l ⁻¹)	COD (mg(O ₂) l ⁻¹)*
IEC A	0.21 ± 0.02	0	700.0	12,178.86 ± 285.55
IEC B	0.21 ± 0.02	0	700.0	12,422.76 ± 149.03
TEV	0.21 ± 0.02	0	714.0	7,455.29 ± 212.44
HIC A	1,206.85 ± 116.80	5,170.97 ± 24.74	-	52.03 ± 8.68
HIC B	1,216.51 ± 96.57	0	-	48.78 ± 15.73
SEC	1,192.10 ± 65.02	0	-	308.94 ± 7.01
IMAC A	1,197.37 ± 72.56	15.12 ± 0.96	1,121.0	3,520.33 ± 172.92
IMAC B	1,197.19 ± 71.83	205.62 ± 6.34	14,010.0	41,463.42 ± 739.79

* The data shows mean values ± standard deviation (n = 6)

The results showed the highest concentration of NH₃ (5,170.97 mg(NH₃) l⁻¹) in the HIC A buffer as it is required in this chromatographic method to increase the hydrophobicity of proteins ensuring binding to the resin [Jacob 1999]. In solution the (NH₄)₂SO₄ salt added is present as ammonium ions (NH₄⁺) resulting in high NH₃ values obtained during the buffer analysis. Ammonia was also detected in IMAC A and IMAC B (15.12 and 205.62 mg(NH₃) l⁻¹, respectively) as both of these buffers contain imidazole. Imidazole (1, 3-diazacyclopenta-2, 4-diene) is synthesised from glyoxal and formaldehyde in ammonia which may still be present at low concentrations in the final imidazole product [Revuelta *et al.* 2011].

Organic nitrogen (N_{org}) was determined theoretically from the chemical structures of the compounds added to each buffer. Each imidazole molecule has two organically linked nitrogen atoms, therefore, the imidazole containing buffers (IMAC A and IMAC B) have high N_{org} amounts. N_{org} is also present in the IEC A, IEC B and TEV buffer which are Tris-based solutions. As Tris (2-amino-2-hydroxymethyl-propane-1,

3-diol) has only one N atom in each molecule the overall N_{org} concentration in these buffers is lower compared to the IMAC buffers. Additionally, TEV contains low concentrations of ethylenediaminetetraacetic acid (EDTA), a chemical compound with two organically linked N atoms.

The COD analysis identified IMAC B as the buffer with the highest COD amount of 41,463.42 $\text{mg}(\text{O}_2) \text{ l}^{-1}$, followed by IEC A and IEC B with 12,178.86 and 12,422.76 $\text{mg}(\text{O}_2) \text{ l}^{-1}$, respectively. The high COD values are due to the presence of imidazole in IMAC B and Tris and maleic acid in IEC A and IEC B. While imidazole is also added to IMAC A, it has at much lower concentration compared to IMAC B which is reflected by the lower COD value of 3,520.33 $\text{mg}(\text{O}_2) \text{ l}^{-1}$. TEV had a COD concentration of 7,455.29 $\text{mg}(\text{O}_2) \text{ l}^{-1}$ due to the Tris salts present in this buffer. The trypsin inhibitor, a protein, in conjunction with the dithiothreitol (DTT) resulted in a COD value of 308.94 $\text{mg}(\text{O}_2) \text{ l}^{-1}$ for the SEC buffer. In HIC A and HIC B the only compound affecting the COD is the trypsin inhibitor, therefore, the COD values are relatively low in both buffers (52.03 and 48.78 $\text{mg}(\text{O}_2) \text{ l}^{-1}$, respectively).

The environmental analysis of all the buffers did indicate that some of them have higher TP concentrations while other buffers had higher N or COD values. It was therefore difficult to determine from just these results, which of the two purification schemes would have a lower overall environmental impact. To allow a clear comparison of the environmental factors between the two purification schemes the findings were normalised to show the results of each analysis per mg of purified β -galactosidase (Table 4.4).

Using the affinity-based purification scheme reduces the TP, NH_3 and COD concentration by 46.2, 99.4 and 61.6 %, respectively. A significant amount of NH_3 is saved by avoiding the use of HIC which requires the use of ammonium sulphate to facilitate protein binding. Other salts may be used instead but one of the most commonly employed salts during HIC is $(\text{NH}_4)_2\text{SO}_4$ [GE Healthcare 2010b]. The drop in TP and COD values is a result of the reduced amount of buffer in the affinity-based purification method per mg of purified product.

Table 4.4 Environmental analysis of the various buffers employed and comparison of the two purification schemes

Environmental analysis	Traditional multi-chromatographic scheme	Affinity-based purification scheme*
TP (mg P per mg purified protein)	309.61	166.64 (-46.2 %)
NH ₃ (mg NH ₃ per mg purified protein)	1,100.64	6.33 (-99.4 %)
N _{org} (mg N _{org} per mg of purified protein)	261.50	461.67 (+76.5 %)
COD (mg O ₂ per mg purified protein)	4,141.78	1,590.02 (-61.6 %)
Water volume (ml H ₂ O per mg of purified protein)	613	154 (-74.9 %)

* Affinity-based purification scheme shows values and in brackets % difference to traditional multi-chromatographic purification scheme.

The table is normalised to show the results per mg of purified recombinant protein.

While TP is an important specific environmental indicator, the determination of the COD, which estimates the organic strength of a solution, is often used to evaluate the total pollutive potential of a waste stream, i.e. the higher the COD value the higher the amount of pollution in the sample [Hammer and Hammer 2004]. Discharging phosphorus into the environment can lead to eutrophication and deterioration of the water quality [Berg *et al.* 2006] and high COD concentrations are toxic to biological life and will affect the aquatic environment [Musa and Ahmad 2010]. Therefore, industrial waste streams containing high TP and/or COD values require treatment prior to release into aquatic bodies. In general, waste streams of industries within municipal limits are discharged to the city's wastewater treatment system after pretreatment, which is necessary if there is a significant difference in strength and characteristics between the industrial and sanitary wastewater. While it may be possible to dilute manufacturing wastes with domestic wastewater if high concentrations of COD are observed, high TP concentrations in waste streams can require treatments, employing chemical or biological removal techniques, at the industrial sites [Forster 2003; Hammer and Hammer 2004]. Therefore, the reductions in TP and COD in the waste streams that arise during large-scale protein purification using an affinity-based strategy can have environmental as well as cost benefits for a company.

Phosphorus is removed from wastewater through incorporation of P into total suspended solids (TSS) and subsequent removal of those solids. This may be achieved by integration of P into either biological solids (microorganisms) or chemical precipitation. Chemical P removal involves the use of divalent cations which form sparingly soluble P precipitates. The most commonly employed ions are calcium, aluminium and iron but polymers in conjunction with alum or lime may also be used [Tchobanoglous *et al.* 2003; Russell 2006]. An alternative process relies upon enhanced biological phosphorus removal (EBPR) which involves so-called luxury P uptake by a specific group of heterotrophic bacteria. These polyphosphate-accumulating organisms are selectively enriched in the bacterial community within the activated sludge [Al-Dasoqi *et al.* 2011; van Haandel and van der Lubbe 2012]. Phosphorus-containing sludge obtained using either method can be landfilled, incinerated or recycled as a fertiliser for land application [Tchobanoglous *et al.* 2003]. The requirement of P removal from wastewater can significantly increase the costs involved in wastewater treatment, because of the added chemical costs and/or the increased quantities of sludge to be disposed of [Paul *et al.* 2001; Shu *et al.* 2006; van Haandel and van der Lubbe 2012].

Due to the high concentration of imidazole used during IMAC, the amount of N_{org} per mg of purified protein is higher by 76.5 % for the affinity-based product isolation scheme. Imidazole is employed during the chromatographic step as it has a high efficiency for replacing polyhistidine tags and interacts with the metal ion, thereby, facilitating elution of bound proteins from the column. Low concentrations of imidazole were added to the wash buffer (IMAC A) to facilitate the removal of weakly bound host cell proteins and increase purity of the target protein. The imidazole concentration selected for elution (in IMAC B) ensures complete removal of all bound proteins avoiding cross-contamination for future runs [Kastner 1999; Petty 2001; GE Healthcare 2010b].

Another major concern can be the extensive use of water during downstream processing. Typical water usage for large-scale chromatography and ultrafiltration/diafiltration can range from 100 to 1,000 l of water per kg of product for each step [Ho 2010]. To minimise or eliminate contaminants in the biopharmaceutical product, the water used during downstream processing is “water

for injection” (WFI) which is generated via extensive purification steps [Walsh 2003b]. WFI is often the most expensive utility employed in a biopharmaceutical plant and adds considerably to overall production costs [Riedewald *et al.* 2012].

During the traditional multi-chromatographic scheme 664 ml of water per mg of purified protein were employed while using the affinity-based purification strategy required only 165 ml per mg purified product. This 75 % reduction in overall water consumption was achieved due to a reduction in the total number of steps from six to four unit operations when using IMAC.

This analysis has shown that utilising an affinity-based scheme for purifying recombinant His₆-tagged proteins results in significant reduction in TP, NH₃, COD and overall water consumption. That indicates that this purification strategy is the better option in terms of environmental impact of downstream processing for large-scale protein production.

4.4.4 Cost analysis for both purification schemes

A cost analysis breakdown of the laboratory-scale purification for both strategies is shown in Table 4.5. A cost reduction of 65 % was achieved when comparing just the chemicals involved during the purification of recombinant β -galactosidase using the affinity-based method. These savings are a result of decreasing the number of steps and therefore buffer usage but also due to the increased yield that was attained when utilising affinity-based chromatography. However, the major expenditure in downstream processing is the chromatographic media and made up 99 % of the overall costs for the laboratory-scale purification. Compared to other purification methods the chromatographic media used for IMAC can involve much higher costs [Stiborova *et al.* 2003] but IMAC has shown to achieve very high yields sometimes over 90 % in laboratory-scale environments [Arnau *et al.* 2006].

The cost analysis showed that the reduction in number of chromatographic steps and the increase in yield resulted in an overall drop in media costs per mg of purified product of 60 % when employing the affinity-based purification scheme for the recombinant β -galactosidase. This is also reflected in the total costs involved in isolating the recombinant protein. This study has shown that it is possible to reduce

the downstream process costs by 60 % when using a His₆-tag and an IMAC purification strategy.

Table 4.5 Cost breakdown of the two purification schemes for the isolation of the recombinant β -galactosidase

Cost breakdown	Traditional multi-chromatographic scheme	Affinity-based purification scheme*
Chemicals (Costs per mg of purified product)	€5.68	€1.99 (-65.0 %)
Chromatographic media (Costs per mg of purified product)	€552.02	€220.00 (-60.1 %)
Total (Costs per mg of purified product)	€557.70	€221.99 (-60.2 %)

* *Affinity-based purification scheme shows values and in brackets % difference to traditional multi-chromatographic purification scheme.*

The table is normalised to show the results per mg of purified recombinant protein.

Large-scale downstream processing requires high production efficiencies as well as low costs of the final product for the process to be commercially viable [Palomares *et al.* 2004]. The affinity-based strategy employed in this study has been shown to fulfil these criteria. The use of IMAC in combination with a polyhistidine tag allows high purification efficiency and, due to the reduction of overall process steps, cost savings can be achieved.

4.4.5 Reduction of environmental impact of affinity-based purification

Minimisation of the environmental impact, in particular phosphorus outputs, was investigated using the affinity-based scheme for purifying the recombinant β -galactosidase. The aim was to reduce the concentration of TP without adversely affecting the purification factor or yield. Section 4.4.3 clearly showed that by employing His₆-tags and IMAC for the purification of recombinant proteins, a reduction in environmental impact in downstream processing can be achieved. However, as part of this study it was evaluated if there is a potential to further reduce or even remove phosphorus from the waste streams that arise during recombinant protein purification.

An initial attempt was made to reduce the TP concentration in waste streams by reducing the amount of P salts added to buffers used during chromatography as outlined in section 4.3.7. Figure 4.8 summarises the chromatograms obtained for four SEC runs with varying concentrations of P salts in the purification buffer. The results clearly show that it is possible to significantly reduce the concentration of P salts in the SEC buffer during this step without affecting the purification profile. The chromatograms show one large enzyme activity peak between fractions 28-33 which corresponds to the chromatograms obtained during the purification of the recombinant product using the unaltered affinity-based scheme. The yields of the SEC step for all chromatography runs with the varying buffers are greater than 100 % (Table 4.6). This was calculated as a percentage of the amount of recombinant β -galactosidase loaded onto the column during each run. The SEC load was partially purified protein in TEV buffer which has a pH of 8.0. The recombinant β -galactosidase loses activity at that pH as previously shown in Figure 2.14 in section 2.4.10. During SEC elution the recombinant protein is transferred into a buffer with a pH value of 6.5 which is the optimum pH for this β -galactosidase enzyme resulting in an increase of yield.

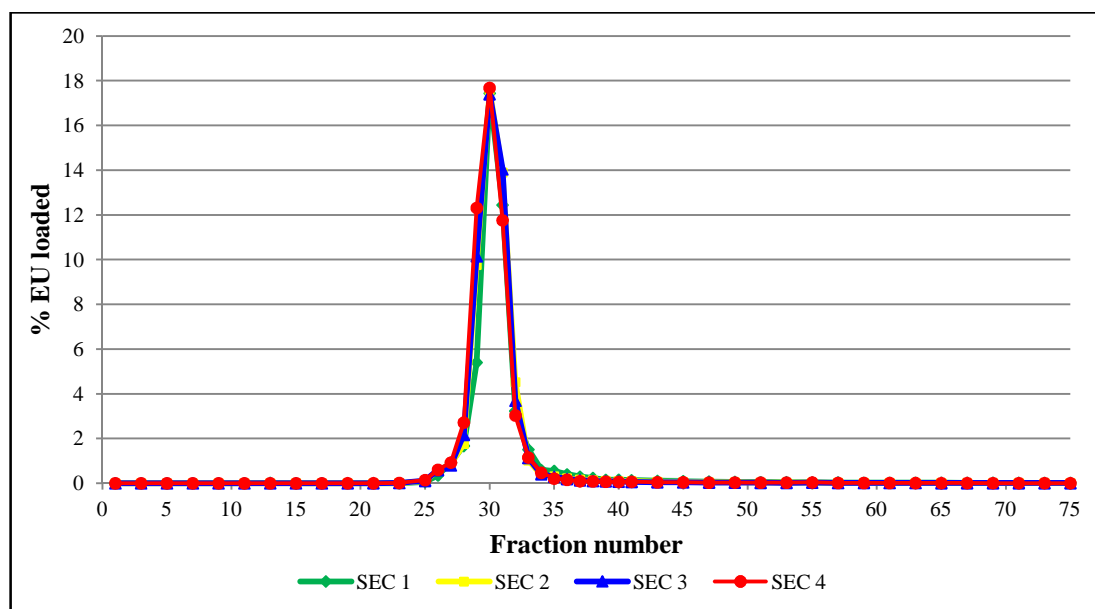


Figure 4.8 Chromatogram showing the results of SEC runs for the recombinant β -galactosidase using buffers with varying phosphate salt concentrations

SEC 1: 50 mM sodium phosphate buffer, SEC 2: 25mM sodium phosphate buffer, SEC 3: 10 mM sodium phosphate buffer, SEC 4: 5 mM sodium phosphate buffer; all buffers also included: 150 mM NaCl, 1 mM DTT and 50 mg ml⁻¹ trypsin inhibitor.

It was shown during the minimisation of TP in fermentation media that reducing the P salt concentration can affect the pH of the solution. Therefore, the SEC peak fractions containing purified β -galactosidase were combined and the pH of each peak tested. The results are tabulated in Table 4.6.

Table 4.6 Summary of results obtained during minimisation of P salts in SEC buffer

Buffer	pH SEC buffer	pH of SEC peak	Yield (%) [*]
50 mM sodium phosphate buffer	6.50	6.49	112
25 mM sodium phosphate buffer	6.50	6.50	121
10 mM sodium phosphate buffer	6.50	6.47	116
5 mM sodium phosphate buffer	6.50	6.39	110

^{*} The yield was calculated in relation to the amount of protein loaded during each run.

The starting pH for all buffers was adjusted to the protein's optimum pH value of 6.5 and reducing the P salts in the SEC buffer up to 50 % resulted in no change of pH in the SEC elution peak. When decreasing the sodium phosphate concentration of the buffer by 80 % (to 10 mM) a minimal drop of the pH value to 6.47 could be observed. However, further decreasing the P salts in the buffer has led to a significant drop of pH of the SEC elution peak (pH 6.39). A change in pH can affect protein structure and stability [Janson 2012] and should therefore be avoided during purification. These findings show that it is possible to reduce the concentrations of P salts in a sodium phosphate buffer by up to 80 % without an adverse affect on the purification profile, yield or pH of the solution during SEC. Such reductions can translate to significant P savings during large-scale protein purification which in turn reduces the environmental impact of the waste streams that are produced during these unit operations.

In addition to minimising the TP concentration in the buffer by reducing the amount of P salts added, alternative buffering compounds that may be used during the purification of the recombinant protein via the affinity-based scheme were investigated for their potential environmental impact (section 4.3.7). The results of these analyses are summarised in Table 4.7.

Table 4.7 Environmental analyses of alternative buffering systems (50 mM concentration) for the use of recombinant protein purification via the affinity-based scheme

Chemical/ buffer Name	Chemical Formula	TP (g(P) l ⁻¹)*	COD (g(O ₂) l ⁻¹)*	Theoretical N _{org} (g(N _{org}) l ⁻¹)
Sodium phosphate	NaH ₂ PO ₄ & Na ₂ HPO ₄	1.18 ± 0.08	n. d.	0
ADA	C ₆ H ₁₀ N ₂ O ₅	0	6.03 ± 0.14	1.40 (2N)
ACES	C ₄ H ₁₀ N ₂ O ₄ S	0	3.97 ± 0.07	1.40 (2N)
BES	C ₆ H ₁₅ NO ₅ S	0	9.51 ± 0.05	0.70 (1N)
Bis-Tris	C ₈ H ₁₉ NO ₅	0	14.41 ± 0.34	0.70 (1N)
Bis-Tris propane	C ₁₁ H ₂₆ N ₂ O ₆	0	15.74 ± 0.53	1.40(2N)
HEPES	C ₈ H ₁₈ N ₂ O ₄ S	0	12.46 ± 0.29	1.40 (2N)
Clark and Lubs	KH ₂ PO ₄ & NaOH	1.31 ± 0.05	n. d.	0
MES	C ₆ H ₁₃ NO ₄ S	0	9.62 ± 0.19	0.70 (1N)
PIPES	C ₈ H ₁₈ N ₂ O ₆ S ₂	0	11.98 ± 0.31	1.40 (2N)
TES	C ₆ H ₁₅ NO ₆ S	0	8.71 ± 0.19	0.70 (1N)
Tricine	C ₆ H ₁₃ NO ₅	0	9.41 ± 0.10	0.70 (1N)
Tris-Acetate	C ₄ H ₁₁ NO ₃ & C ₂ H ₃ O ₂ ⁻	0	10.01 ± 0.19	0.70 (1N)

* The data shows mean values ± standard deviation (n = 6)

n. d.: none detected

It is clear from the chemical formula of the compounds in the buffer systems as well as the direct TP determination that phosphorus could only be detected in buffering solutions containing P salts (NaH₂PO₄, Na₂HPO₄ or KH₂PO₄). The sodium phosphate buffer originally used during the affinity-based purification and the Clark and Lubs buffer tested contained TP concentrations of 1.18 and 1.31 g(TP) l⁻¹. The Clark and Lubs buffer was investigated to determine if using potassium phosphate salts instead of sodium phosphate influences the overall TP concentration of the solution. The organic N concentration was 0, 0.7 or 1.4 g(N_{org}) l⁻¹ depending on if there is zero, one or two organically linked N atoms present in the compound. The COD values varied significantly with the highest being 15.74 g(O₂) l⁻¹ for the buffer containing Bis-Tris propane and the lowest for P salt-based buffers which had no detectable COD. All of the alternative buffering salts not containing any P showed COD values which would be considered high strength (≥ 0.8 g(O₂) l⁻¹) and wastewater treatment would be necessary prior to discharge [Tchobanoglous *et al.*

2003]. The ACES buffer had the lowest COD value with $3.97 \text{ g(O}_2\text{)} \text{ l}^{-1}$ in addition to no P salts and was selected for purification trials via the affinity-based scheme.

The model recombinant protein (His₆-tagged β -galactosidase) was purified using the affinity-based strategy as described in section 4.3.7 during which the IMAC and SEC sodium phosphate buffers were replaced with the alternative buffers (ACES). Figure 4.9 is a typical chromatogram for the IMAC step when using the ACES buffer during product isolation.

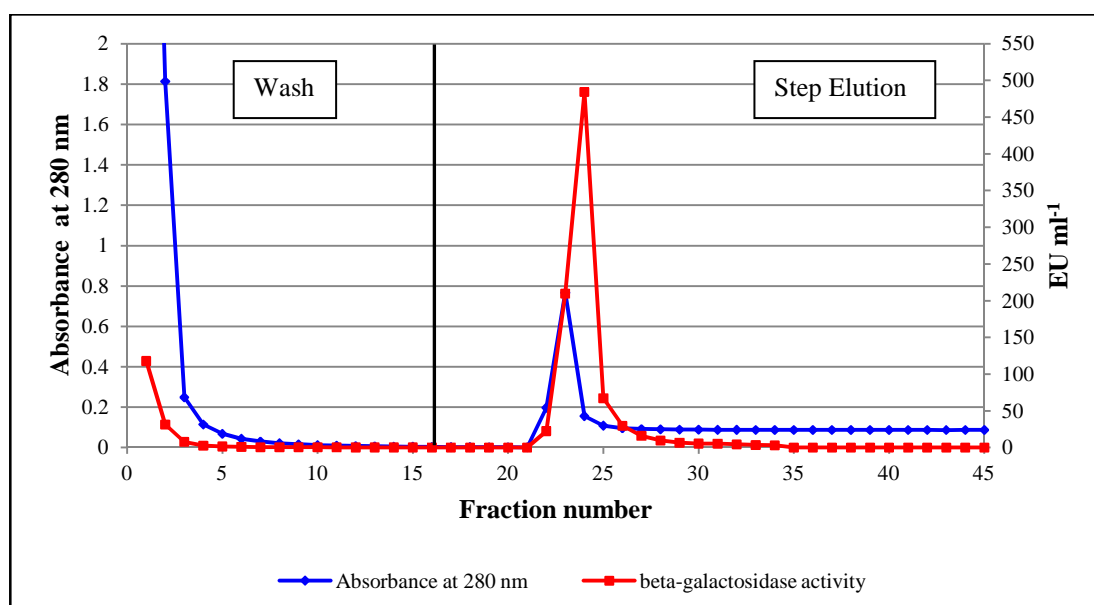


Figure 4.9 Chromatogram for the purification of recombinant β -galactosidase with IMAC using ACES salts as the alternative buffering system

The IMAC chromatogram obtained during the affinity-based purification schemes using the sodium phosphate (Figure 4.5) and ACES (Figure 4.9) buffering systems are comparable. This chromatographic step with either buffer facilitates the removal of a significant quantity of contaminating proteins during the wash step which is shown by the first protein peak (fractions 1-8) in the chromatogram. A small amount of β -galactosidase is also lost during the column washing, however, the majority of recombinant protein is eluted and collected during the step elution as indicated by the enzyme activity peak in the elution fractions (22-29).

The partially purified protein (IMAC peak) was concentrated 8.9 fold and the buffer exchanged via UF/DF for subsequent His₆-tag cleavage. To ensure complete removal of the tag, the cleavage was monitored using SDS-PAGE and Western blot.

Following digestion, the protein mixture was loaded onto the SEC column, which allowed the removal of any remaining contaminants as well as higher order oligomers of the protein of interest. The results of a typical SEC run obtained during the purification of the recombinant product via the affinity-based scheme employing an ACES buffer are illustrated in Figure 4.10. Seven protein peaks can be observed in the chromatogram, two of which overlap with the two enzyme activity peaks.

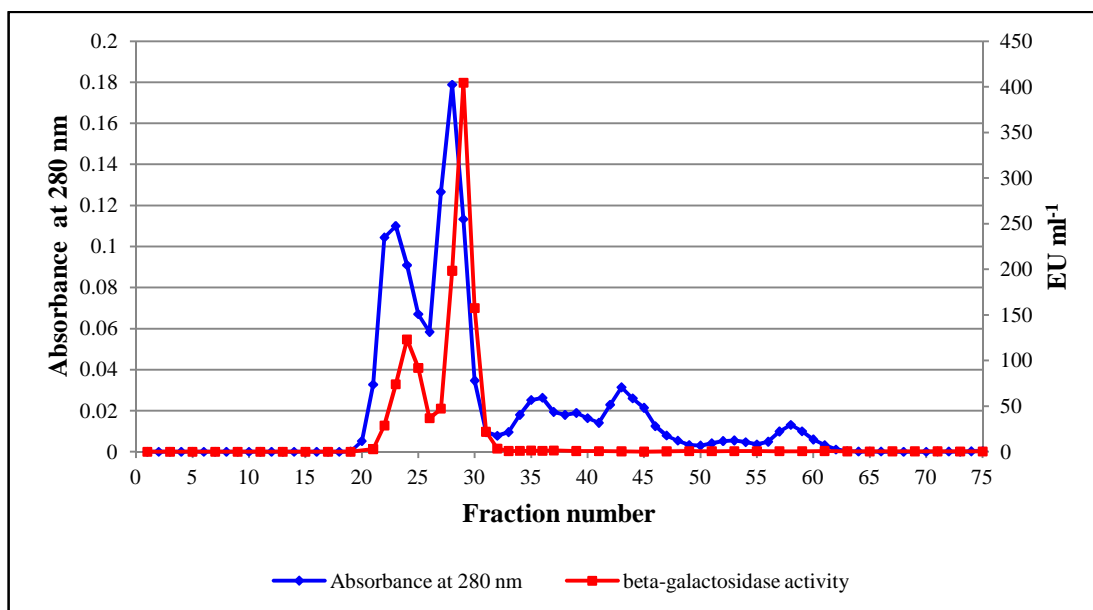


Figure 4.10 Chromatogram for the purification of recombinant β -galactosidase with SEC using ACES salts as the alternative buffering system

The first protein and β -galactosidase activity peak (fractions 22-26) corresponds to previously observed higher order oligomeric forms of the recombinant product while the second peak (fractions 27-33) coincides with the protein of interest, i.e. recombinant β -galactosidase in tetrameric, functional form. A similar chromatogram for the SEC step was obtained when using the sodium phosphate buffers during the affinity-based scheme. This suggests that the purification profile is not altered by replacing the sodium phosphate buffers with ACES buffers during the purification of the recombinant protein via the affinity-based strategy.

The results of the affinity-based purification using the ACES buffers for the recombinant β -galactosidase are shown in Table 4.8. The yield (44.2 %) and purification factor (2.83) achieved are comparable to those obtained when using the

sodium phosphate buffers with that affinity-based purification scheme (yield 34.6 % and purification factor 3.02; Table 4.2).

Table 4.8 Purification table for recombinant β -galactosidase purified using the affinity-based chromatography scheme with ACES buffers

Step	Total volume (ml)	Total protein (mg)	Total activity (EU)	Specific activity (EU mg ⁻¹)	Yield (%)	Purification factor
Crude extract	3.1	19.75	3,511.2	177.8	100	1
IMAC peak	24.0	6.71	1,555.7	231.7	44.3	1.3
UF/DF	2.7	5.58	1,515.9	271.7	43.2	1.53
TEV-cleaved protein	2.7	5.64	1,812.2	321.3	51.6	1.81
SEC peak	14	3.09	1,552.4	502.4	44.2	2.83

Crude and purified protein samples of recombinant β -galactosidase isolated using the affinity-based scheme with an ACES buffering system were analysed by SDS- and native PAGE according to section 2.3.8. The SDS-PAGE gel (Figure 4.11 **A**) showed that the purified enzyme migrated as a single band, indicating purification to homogeneity. The molecular mass of the protein band was estimated with respect to the protein standards of known molecular masses to be 116.6 kDa which corresponds to the theoretical molecular mass of the recombinant β -galactosidase monomer (without His₆-tag) as determined according to section 2.3.2. Native PAGE analysis (Figure 4.11 **B**) showed a single band illustrating that the β -galactosidase was purified to homogeneity and the higher order oligomers were removed by the SEC. Enzymatic activity of the purified recombinant product was confirmed through the zymogram and native PAGE overlay technique (Figure 4.11 **B**)

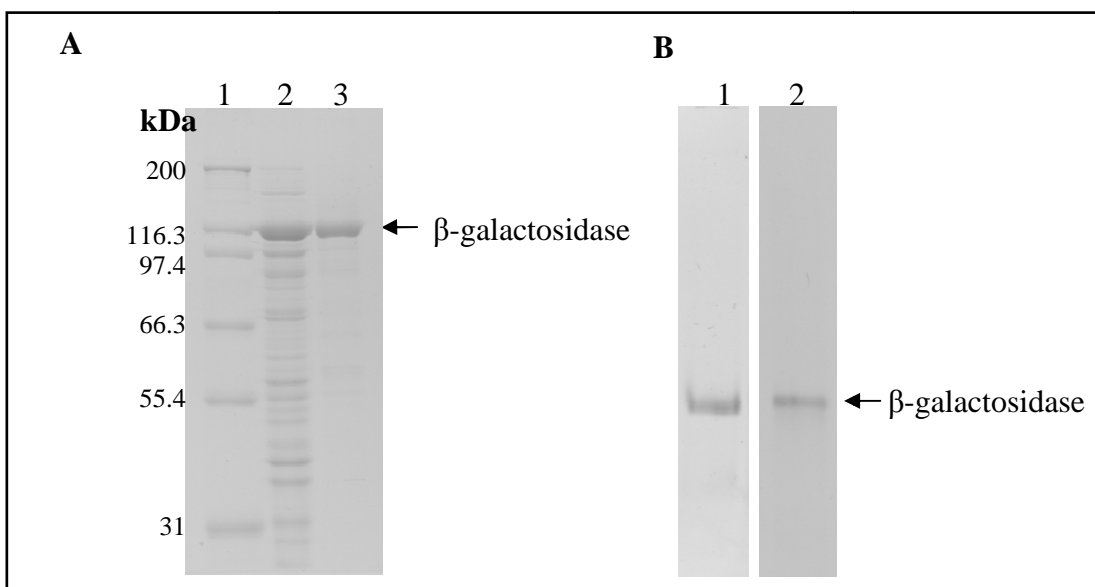


Figure 4.11 Electrophoretic analysis of recombinant β -galactosidase purified using the affinity-based chromatography scheme with ACES buffering system

A

SDS-PAGE analysis stained for protein – Lane 1: Protein Standard Mark 12 (Life Technologies); Lane 2: crude protein preparation; Lane 3: purified protein.

B

Native PAGE analysis of purified enzyme stained for protein (Lane 1) and stained for activity (Lane 2).

These findings showed that it is possible to replace the P salts in buffering systems used during affinity-based protein purification with alternative salts such as ACES without negatively affecting the isolation of recombinant His₆-tagged products.

The ACES salt is more expensive than sodium phosphate which results in an increase in the cost of chemical raw materials employed by 57 % compared to affinity-based purification using sodium phosphate buffers. However, the costs of the chemicals still only make up 2 % of the total costs for downstream processing and in combination with the cost savings that could potentially be observed during wastewater treatment it is likely that the use of ACES as an alternative buffering system during affinity-based purification is economically feasible.

In addition, ACES is not considered to be a hazardous substance or mixture according to Regulation (EC) No. 1272/2008 and it is not classified as dangerous as per Directive 67/548/EEC making it a suitable buffer for processing proteins destined for biopharmaceutical use [Sigma-Aldrich 2013a].

4.5 Conclusion

Industrial focus has moved towards minimisation of the environmental impacts during process design, development and manufacturing due to increased environmental awareness and regulations as well as escalating environmental control costs. This research was aimed at comparing two downstream processing schemes in terms of their associated environmental impact and integrating a pollution minimisation strategy for large-scale recombinant protein purification.

The initial part of this investigation showed that the model protein (His₆-tagged β -galactosidase) can be purified by employing either traditional multi-chromatographic methods or an affinity-based scheme utilising the polyhistidine tag. While both isolation methods showed comparable final purification, using the affinity-based strategy resulted in a much higher product yield due to IMAC's high specificity for the His₆-tag. Recombinant proteins are amongst the most expensive of all drugs available [Farid 2007] and to ensure initial and continued commercial viability of biopharmaceutical manufacturing high downstream processing yields play a vital role in making the process cost effective.

In addition to increased product recovery, the use of affinity-based protein recovery also reduces the number of steps and overall buffer consumption which in turn decreases the amount of environmentally harmful chemicals in the waste streams produced during downstream processing. The combined environmental benefits and economic savings, without any adverse product consequences, identified genuine industrial potential for the affinity-based purification scheme.

The use of affinity tags is widespread in several areas of research and since most biopharmaceuticals are produced via recombinant protein technology, the concept of including purification tags such as the His₆-tag seems obvious. These tags would simplify the separation process as well as facilitate cost savings giving the increased yield, reduced production costs and wastewater treatment savings [Arnau *et al.* 2006; Cramer and Holstein 2011]. However, a therapeutic protein designed for human administration requires to have a structure that is as similar to the authentic native product as possible, therefore, the presence of an affinity tag needs to be avoided

[Lauritzen *et al.* 2013]. Standard methods for complete tag removal can make the affinity-based purification process expensive, complex and problematic [Cramer and Holstein 2011]. Nonetheless, in recent years advances are being made employing new proteolytic-based approaches for more efficient and precise cleavage of the affinity-tags (e.g. SUMO and TAGZyme) which are showing potential for commercial application [Lauritzen *et al.* 2013]. Another emerging alternative is the use of self-cleaving purification tags, many of which are based on inteins, as they do not require enzymatic cleavage. Additionally, this technology enables affinity-purification and cleavage in a single step. This approach has significant potential for industrial applications, however, its implementation for biopharmaceuticals is still at an early stage [Fong *et al.* 2010; Cramer and Holstein 2011; Li 2011].

This study has also shown that by decreasing or replacing P salts in the buffers used during downstream processing, the TP concentrations in waste streams can be further reduced or even eliminated.

**Chapter Five – Mass balance simulation of phosphorus
wastes arising from recombinant protein production
using *E. coli***

5.1 Introduction

The development of a biopharmaceutical process typically progresses through the following stages: (a) bench-scale realisation of the process mechanisms, (b) process development and materials selection at the pilot plant-scale and (c) process engineering and deployment at the manufacturing plant [Linninger *et al.* 1995]. This chapter focuses on the laboratory-scale simulations of recombinant protein production in *E. coli* (process development stage) by employing the designed model (chapter two) and the upstream and downstream process parameters as outlined in the previous chapters (chapter three and four, respectively). This allows an early understanding of the future production process and the simulation at laboratory-scale can assist in decision making during the process development.

Traditionally, environmental impact concerns were considered only at the later stages of process development and generally involved end-of-pipe technologies to reduce the environmental burden [Li *et al.* 2009]. In recent years however, waste management has focused on waste prevention or reduction using e.g. cleaner technologies as well as waste minimisation via computer aided design [New *et al.* 2000]. Research interests now include alternative approaches to implementation of pollution control in industries.

The simulation of recombinant protein manufacturing is beneficial as it quantitatively establishes and characterises the waste streams generated during production at the initial process development phase. This may assist in process design most specifically in terms of reducing the environmental impact of the overall production run, creating more environmentally friendly and efficient processes while also generating smaller quantities of harmful wastes [Linninger *et al.* 1995].

In this chapter, three processes showing the production of recombinant β -galactosidase were simulated in full at laboratory-scale:

1. Fermentation of model *E. coli* strain with TB media and downstream processing using the traditional multi-chromatographic purification scheme
2. Fermentation of model *E. coli* with TB media and downstream processing employing the affinity-based purification strategy

3. Fermentation of model *E. coli* with 70 % P-reduced TB media and downstream processing via affinity-based purification using the alternative buffering system (ACES salts)

The specific aim during this element of the research was to generate total phosphorus mass balances for each of the processes as well as to assess and compare the TP emissions between the different production designs. This will allow an optimisation of the process in terms of environmental impact during the early development phase.

5.2 Materials

Materials, reagents and microbial strains were obtained as described in section 2.2 and 4.2 unless otherwise stated.

5.3 Methods

5.3.1 Mass balance

The raw materials utilised as well as wastes, emissions and by-products arising from the recombinant protein production were identified by analysis of the materials and methods of the selected processes. The input and output streams were illustrated with flow diagrams using Microsoft[®] Office Word[®] 2007. Subsequently, the inputs and outputs for the mass balances were quantified using measurements and experimental analyses to evaluate the pollution load from the operation (section 5.3.4).

5.3.2 Upstream processing

TB and P-reduced media were prepared as outlined in sections 2.2.4 and 3.4.2, respectively. The model *E. coli* strain MC1061 carrying pProEx-HTb_ *lacZ*-HT was grown and the protein expressed using the optimised conditions as detailed in section 2.4.8. The cell pellets were harvested via centrifugation at 7,500 rpm, 4 °C and the cell pellets lysed via sonication using 30 s bursts for 20 min with intermittent cooling. Subsequent centrifugation at 12,000 rpm, 4 °C for 20 min yielded the soluble crude extract which contains the functional fraction of the recombinant product.

5.3.3 Downstream processing

The recombinant β -galactosidase was purified from the crude protein solutions via the traditional multi-chromatographic scheme, the affinity-based purification method and the affinity-based strategy with the alternative buffering system as described in sections 4.3.2, 4.3.3 and 4.3.7, respectively.

5.3.4 Environmental analyses of waste streams

The waste streams for the IEC, HIC and IMAC steps included the 10 CV of pre-equilibration buffers as well as the wastes obtained during each run. Total

phosphorus analysis was undertaken on all waste streams generated during the recombinant protein production by reducing the phosphate fractions to orthophosphate which was determined colorimetrically using the stannous chloride method. The complete methodology is outlined in section 3.3.7.

In addition, the polluting quality of all the waste streams from the recombinant protein manufacturing process was determined by chemical oxygen demand (COD) analysis. Using a mixture of chromic and sulphuric acids the organic matter in the samples was oxidised. The oxygen consumed was measured against standards at 600 nm with a spectrophotometer. Full details are described in section 4.3.4.

5.3.5 SuperPro Designer[®] 6.0

The specific input and output volumes as well as chemical and biological wastes arising from recombinant protein production were diagrammatically represented using the modelling software SuperPro Designer[®] 6.0 (Intelligen, Europe Inc., Thessaloniki, Greece).

5.4 Results and Discussion

5.4.1 Waste streams generated during the production of the recombinant model protein

The modelling software SuperPro Designer[®] 6.0, which has been used for the simulation of a variety of processes (e.g. [Biwer *et al.* 2005a; Biwer *et al.* 2005b; Ramírez *et al.* 2009; Kuhn *et al.* 2010; Marchetti 2011; Vázquez and Rodríguez 2011; Brunet *et al.* 2012; Taras and Woinaroschy 2012]), was employed during this study to establish mass balances for a specific pollutant (phosphorus) emitted during recombinant protein manufacturing. The raw material and chemical reagent inputs and waste outputs of the process using the different purification systems were determined and assessed in detail while factors such as energy, water usage etc. were excluded.

5.4.1.1 Content of waste streams produced when employing the traditional multi-chromatographic purification scheme

Figure 5.1 is a schematic flow diagram of the recombinant protein production employing the traditional multi-chromatographic purification scheme as described in section 4.3.2 and portrays the aqueous volumes and solid waste mass arising at each step of the laboratory-scale production process. The flow schematic was designed using SuperPro Designer[®] 6.0 as outlined in section 5.3.5 and indicates that with this purification scheme six aqueous waste streams (W_1 - W_6) and one solid waste (SW) arise. A summary of the chemical and biological waste components generated during recombinant β -galactosidase production via traditional multi-chromatographic purifications is presented in Table 5.1.

The first aqueous waste stream (W_1) emitted from the manufacturing process is at the end of fermentation step and contains unused media constituents, including P-containing compounds (section 3.4.2), antibiotic traces as well as cellular debris, proteins and potentially some residual suspended bacteria. Since the spent fermentation media (W_1) may contain live suspended cells, best practice advises that this waste stream is biologically inactivated prior to release into wastewater treatment plants (WWTP).

Table 5.1 Summary of the waste constituents present in the aqueous wastes 1-6 (W_1 - W_6) and in the solid waste (SW) generated during the recombinant β -galactosidase production using the traditional multi-chromatographic purification scheme

Waste streams	Waste constituents	
	Chemical	Biological
W_1	Water, unused organic components containing N and P, antibiotic traces	Cellular debris, suspended cells, nucleic acids, proteins
SW	Water, salt traces	Cellular debris, denatured proteins
W_2	Water, Tris, maleic acid, NaOH, NaCl	Proteins, nucleic acids, lipopolysaccharides
W_3	Water, Tris, maleic acid, NaOH, NaCl, Tris-Cl, EDTA	Proteins
W_4	Water, sodium phosphate, $(NH_4)_2SO_4$	Proteins
W_5	Water, sodium phosphate	Proteins
W_6	Water, sodium phosphate, NaCl, DTT	Proteins

W_1 : aqueous waste one, W_2 : aqueous waste two, W_3 : aqueous waste three, W_4 : aqueous waste four, W_5 : aqueous waste five, W_6 : aqueous waste six, SW: solid waste; DTT: dithiothreitol, EDTA: 2, 2', 2'', 2'''-(ethane-1, 2-diyl dinitrilo)tetraacetic acid

This is carried out via thermal inactivation such as autoclaving or chemical inactivation involving exposure to extreme pH shifts by alternating addition of acid and base causing cell lysis [D'Souza and Killedar 2008; Behme 2009].

The subsequent waste emitted was the solid waste (SW) generated from the cell lysis step which consists mainly of cellular debris and denatured proteins with trace amounts of nucleic acids, some salts and water. Approximately 1 ml of production volume was lost, likely removed as waste with the cell pellet. Inactivation of the biological components within the pellet is generally carried out via autoclaving prior to disposal to landfill or incineration [Behme 2009]. The solid waste mass was not investigated further because this research focused on analysing liquid waste streams as these are directly discharged to WWTP or to the environment. However, it can be assumed that this solid waste contains a certain amount of nitrogen and phosphorus due to the presence of molecules such as amino acids and membrane lipids.

The remaining waste streams arise during the purification of the recombinant product and consists of mainly water with buffering salts as well as contaminating proteins, small amounts of lipopolysaccharides and nucleic acids.

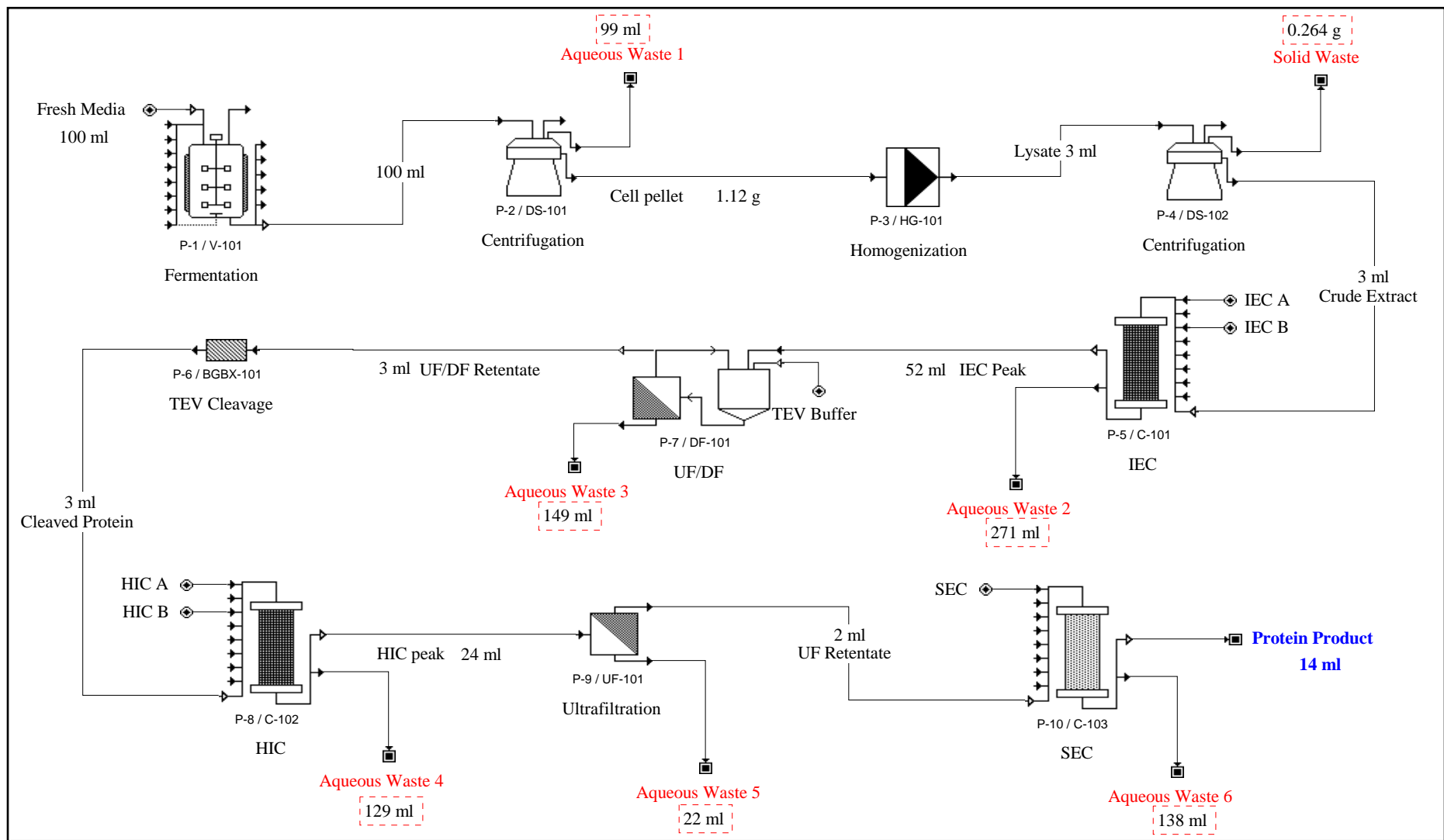


Figure 5.1 Flow schematic generated using SuperPro Desinger[®] 6.0 highlighting the inputs and outputs of the recombinant β -galactosidase production process employing the traditional multi-chromatographic purification method

Production streams and waste output volumes (ml) and weights (g) were determined experimentally at laboratory-scale as described in sections 5.3.2 and 5.3.3. The waste volumes/masses are surrounded by a red frame and the product is highlighted in blue.

The second aqueous waste stream (W_2) is produced during the IEC step and its chemical components are Tris, maleic acid, NaOH and NaCl in water. These constituents influence the oxygen consumption and can cause high COD loadings on the WWTP. There may also be some trace amounts of P due to the presence of nucleic acids and lipopolysaccharides as well as carryover from the previous steps. Aqueous waste stream three (W_3) arises from the UF/DF step and contains salts present in the IEC buffers (IEC A and IEC B), i.e. Tris, maleic acid, NaCl and NaOH, as well as the salts of the TEV buffer (Tris and EDTA). Again, these salts will affect the COD value of the waste stream. The biological compounds that can be found in W_3 include proteins and some trace amounts of nucleic acids and lipopolysaccharides. Liquid waste stream four (W_4) comprises of sodium phosphate salts, ammonium sulphate and proteins and arises during HIC. The high concentration of P present in this waste stream may require pretreatment prior to release to a WWTP. These P removal pretreatments involve physical (e.g. microfiltration), chemical (e.g. precipitation with aluminium or iron coagulates) or biological methods (microbial uptake) [Hammer and Hammer 2004; Zhao *et al.* 2012]. The next waste stream (W_5) is generated during the ultrafiltration step and consists mainly of the water, HIC buffering salts and some trace amount of proteins. The last aqueous waste stream (W_6) produced during this manufacturing process is at the end of SEC and consists of sodium phosphate salts, NaCl, DTT and proteins. As a result of the high concentrations of P present in waste streams W_5 and W_6 pretreatment may also be necessary prior to discharge to WWTP.

Overall, it could be established from Figure 5.1 that the largest volume of waste is emitted from W_2 , followed by W_3 , W_6 , W_4 , W_1 and W_5 . However, the liquid waste streams generated during protein purification (W_2 - W_6) are mostly dilute salts compared to W_1 which has a high nutrient concentration and it is likely that this waste stream is the most polluting. To assess this further a mass balance estimate for TP was carried out (section 5.4.2.1).

5.4.1.2 Content of waste streams produced when employing affinity-based purification scheme

Figure 5.2 depicts the recombinant β -galactosidase production employing the affinity-based protein purification method as outlined in section 4.3.3 and shows the

aqueous volumes and solid waste mass that arise at each step of a laboratory-scale process. SuperPro Designer[®] 6.0 was used to design this flow schematic according to section 5.3.5. It clearly indicates that employing this purification strategy results in only four aqueous waste streams (W_1 - W_4) and one solid waste (SW). Table 5.2 is a summary of the chemical and biological waste components found in each waste stream produced during the production via the affinity-based purification. The first aqueous waste stream (W_1) and the solid waste (SW) are the same as for the manufacturing process employing the traditional multi-chromatographic purification as only the steps following cell lysis were altered. These wastes will require the same treatment, i.e. biological inactivation, prior to release.

Table 5.2 Summary of the waste constituents present in the aqueous wastes 1-4 (W_1 - W_4) and in the solid waste (SW) generated during the recombinant β -galactosidase production using the affinity-based purification scheme

Waste streams	Waste constituents	
	Chemical	Biological
W_1	Water, unused organic components containing N and P, antibiotic traces	Cellular debris, suspended cells, proteins
SW	Water, salt traces	Cellular debris, denatured proteins
W_2	Water, sodium phosphate, NaCl, imidazole	Proteins, nucleic acids, lipopolysaccharides (LPS)
W_3	Water, sodium phosphate, NaCl, imidazole, Tris, EDTA	Proteins
W_4	Water, sodium phosphate, NaCl, DTT	Proteins

W_1 : aqueous waste one, W_2 : aqueous waste two, W_3 : aqueous waste three, W_4 : aqueous waste four, SW: solid waste; DTT: dithiothreitol; EDTA: 2, 2', 2'', 2'''-(ethane-1, 2-diyl)dinitrilo)tetraacetic acid

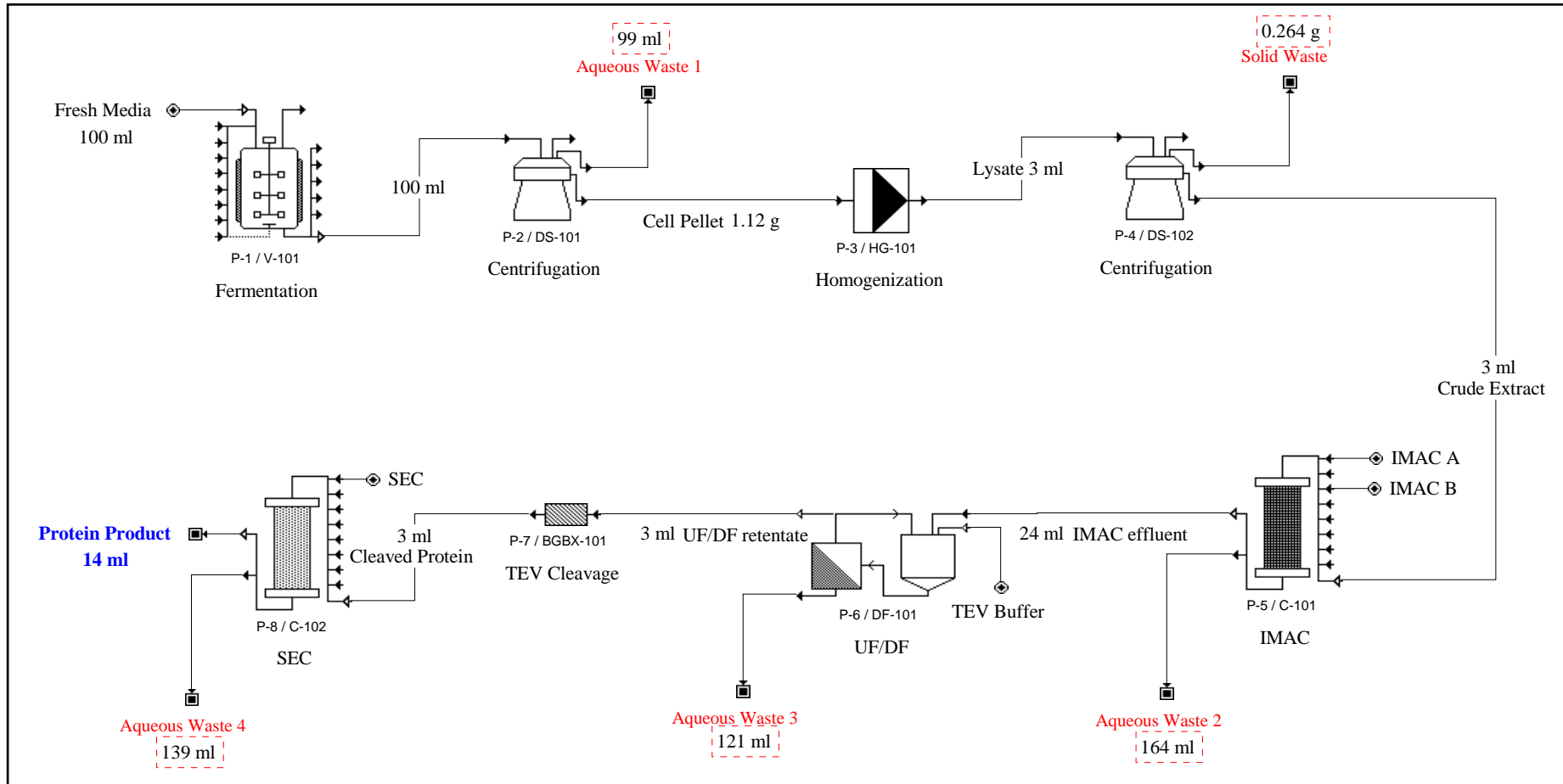


Figure 5.2 Flow schematic generated using SuperPro Designer[®] 6.0 highlighting the inputs and outputs of the recombinant β -galactosidase production process employing the affinity-based purification method

Production streams and waste output volumes (ml) and weights (g) were determined experimentally at laboratory-scale as described in sections 5.3.2 and 5.3.3. The waste volumes/masses are surrounded by a red frame and the product is highlighted in blue.

Aqueous waste stream two (W_2) is emitted after the IMAC step and contains sodium phosphate salts, NaCl and imidazole as well as proteins, nucleic acids and lipopolysaccharides in large quantities of water. UF/DF produces another aqueous waste (W_3) which consists of sodium phosphate, NaCl, imidazole, Tris, EDTA and biological constituents such as proteins and very small amounts of nucleic acids and lipopolysaccharides. The chemicals are a mixture of the buffers employed during IMAC and the TEV buffer for the His₆-tag cleavage. The last waste stream (W_4) is generated during SEC and it includes the chemical components sodium phosphate, NaCl and DTT as well as proteins which are a biological constituent. Due to the high concentration of P in the aqueous waste streams W_2 - W_4 pretreatment may be required prior to discharge to a WWTP.

Employing the affinity-based purification strategy resulted in the largest volume of waste emitted from W_2 , followed by W_4 , W_3 , and W_1 . The liquid wastes produced during the β -galactosidase purification (W_2 - W_4) consist of P salts diluted in water, while W_1 contains a high nutrient concentration from the unused media ingredients. A mass balance estimate for TP was carried out to further evaluate the P pollution potential of each waste stream (section 5.4.2.2). The findings also indicate that employing an affinity-based purification scheme as outlined here can reduce the overall wastes produced during downstream processing which in turn results in reduced wastewater treatment costs.

5.4.1.3 Content of waste streams produced when employing the P-reduced fermentation media and the affinity-based purification scheme with the alternative buffering system

The steps involved in the production of the recombinant product are the same as shown in Figure 5.2, however, the fermentation media employed is the 70 % P-reduced TB media (section 3.4.2) and the buffer for the IMAC and SEC step were altered as outlined in section 4.4.5, i.e. ACES salts were used in place of sodium phosphate. The laboratory-scale operation was carried out as described in section 4.3.7 and since the manufacturing process was not changed, the number and volumes of the aqueous waste streams as well as the mass of the solid waste is the same as shown in Figure 5.2. The chemical and biological compounds found in the wastes that arise during the manufacturing process with the P-reduced media and alternative buffers are summarised in Table 5.3.

Table 5.3 Summary of the waste constituents present in the aqueous wastes 1-4 (W_1 - W_4) and in the solid waste (SW) generated during the recombinant β -galactosidase production using the P-reduced fermentation media and the affinity-based purification scheme with the alternative buffering system

Waste streams	Waste constituents	
	Chemical	Biological
W_1	Water, unused organic components containing N and P, antibiotic traces	Cellular debris, suspended cells, proteins
SW	Water, salt traces	Cellular debris, denatured proteins
W_2	Water, ACES, NaCl, imidazole	Proteins, nucleic acids, lipopolysaccharides (LPS)
W_3	Water, ACES, NaCl, imidazole, Tris, EDTA	Proteins, nucleic acids and LPS
W_4	Water, ACES, NaCl, DTT	Proteins

W_1 : aqueous waste one, W_2 : aqueous waste two, W_3 : aqueous waste three, W_4 : aqueous waste four, SW: solid waste; DTT: dithiothreitol, EDTA: 2, 2', 2'', 2'''-(ethane-1, 2-diylidinitrilo)tetraacetic acid

The changes made to the fresh media for the fermentation step will result in a reduced P concentration in the spent media (W_1). While there are still unused organic compounds containing N and P, the overall amount of TP remaining in the first aqueous waste stream (W_1) is significantly reduced. The solid waste (SW) which arises during cell lysis did not notably change in its composition and still contains mainly cellular debris, denatured proteins, small quantities of water and some salts. Again, W_1 and SW will need biological inactivation treatments prior to release. Liquid waste stream two (W_2) which is produced during IMAC includes high volumes of water with ACES, the alternative buffering salt, NaCl and imidazole as well as the biological components: proteins, nucleic acids and lipopolysaccharides. The chemical composition of the third waste stream (W_3), i.e. waste generated during UF/DF, includes ACES, NaCl, imidazole, Tris and EDTA which is a combination of the alternative IMAC buffers and the TEV buffer. The biological waste in W_3 is the same as described earlier and consists of some proteins with trace amounts of nucleic acids and lipopolysaccharides. The last aqueous waste stream (W_4) observed during the affinity-based purification scheme arises during SEC and for this process simulation contains ACES salts, NaCl, DTT and proteins. As a result of the change in buffering systems from sodium phosphate to ACES salts during IMAC and SEC, the liquid waste streams that arise during downstream processing (W_2 - W_4) no longer

contain large amounts of P. There may be a small concentration of P present due to membrane lipids or carryover from the fermentation media in addition to small quantities from the trypsin inhibitor which was added to all buffers. As outlined in section 4.4.3, the trypsin inhibitor employed may contain up to 10 % sodium phosphate resulting in small amounts of TP in the waste streams.

The largest volume of waste is still emitted from liquid waste stream two (W_2), followed by W_4 , W_3 , and W_1 . As the aqueous wastes produced during the product purification steps (W_2 - W_4) no longer contain sodium phosphate salts the TP concentration in these wastes is significantly reduced and it is likely that the highest P pollution potential is found in W_1 . Mass balance estimates for TP were carried out using the affinity-based purification strategy with the alternative buffering system and the 70 % P-reduced TB media (section 5.4.2.3).

5.4.2 Generation of mass balance for TP emitted throughout the recombinant protein production

Detailed material flow diagrams for total phosphorus were constructed (with experimentally determined values as per section 5.3.4) using SuperPro Designer[®] 6.0 as outlined in section 5.3.5. This allows modelling, evaluation and optimisation of integrated processes which aids in the determination of the specific TP content of the waste streams by providing a clear visualisation of the unit operations and comparison between process steps. In addition, mass balances can highlight potential areas for process improvements or pollution prevention. Phosphorus was selected as this particular environmental parameter is considered to be a major pollutant and is routinely monitored in waste streams [Carty *et al.* 1997; Valsami-Jones 2004]. Mass balances compare the quantities of material entering and leaving a process and are governed by the law of conservation of mass which states that the sum of weights of plant inputs and outputs must be equal (section 1.7.1). Using mass balances enables the tracing of materials and/or pollutant releases at each stage during production to the environmental medium. Ideally, they are designed for specific constituents entering and exiting a process and should be developed for individual units of processes [Dupont *et al.* 2000]. The solid wastes that could be observed in the schematic flowsheets (Figure 5.1 and 5.2) were not analysed further. The focus of this chapter is aimed at the TP concentration in the aqueous wastes generated

because these can directly impact on the wastewater treatment and water courses if released untreated during large-scale manufacturing.

5.4.2.1 Total phosphorus mass balance for recombinant protein production using traditional multi-chromatographic purification

Waste stream and product stream samples obtained during the manufacturing of the recombinant enzyme using the traditional multi-chromatographic purification scheme were analysed for TP concentrations as outlined in section 5.3.4. The results of this P mass balance for the laboratory-scale production is visualised in Figure 5.3. The largest quantity of TP (213.1 mg) exits the process in the spent media (aqueous waste stream one, W_1) followed by waste streams six (W_6), four (W_4), five (W_5), two (W_2), and three (W_3) with 165.0 mg, 152.9 mg, 25.6 mg, 1.62 mg and 0.032 mg of TP per production run, respectively. The TP concentration in W_1 is the product of the P-comprising nutrients found in the spent fermentation media with small amounts of leaked P-containing cellular components. The high quantity of P found in W_4 - W_6 is a result of the P salts used in the buffers for the purification as outlined in sections 4.4.3 and 5.4.1.1. Typical TP concentration in urban wastewater ranges between 4-12 mg(P) l^{-1} , therefore, the aqueous waste streams W_1 and W_4 - W_6 are likely to require pretreatment prior to discharge to local wastewater treatment system due to the difference in P characteristics. Depending on the company's IPPC license, W_2 with a total P output of 1.62 mg per production run which translates to 8.41 mg(P) l^{-1} may not require pretreatment prior to release to the WWTP. An IPPC license for the industrial activity covers permit conditions such as emission limit values (ELVs) which are based on best available techniques (BAT) for each facility. License authorities in conjunction with the EU IPPC bureau help to determine BAT and set specific ELVs for the company [EC 2012].

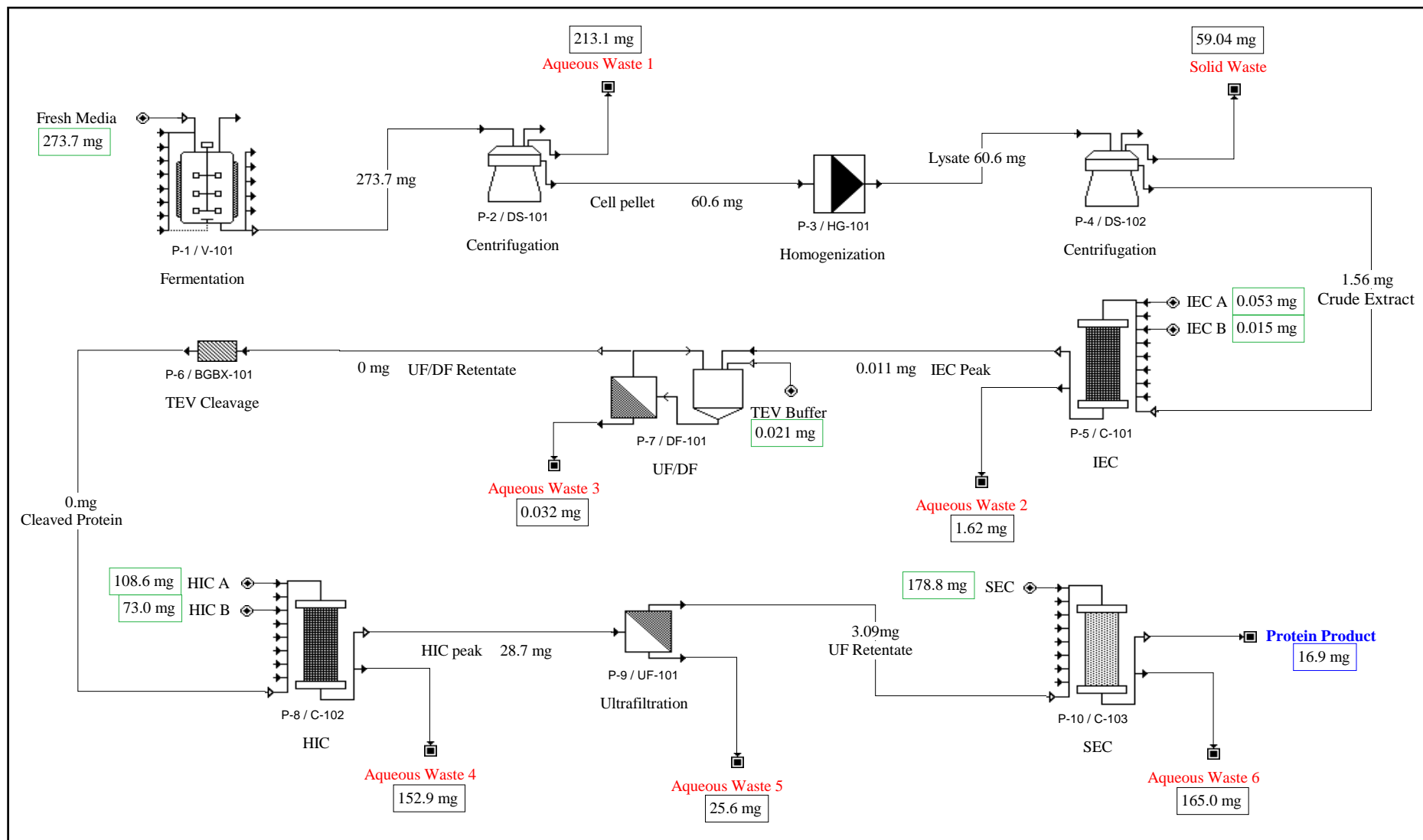


Figure 5.3 Flowsheet detailing TP mass balance for the recombinant protein production employing the traditional multi-chromatographic purification scheme

Results present TP in mg TP waste volume per production run at laboratory-scale as outlined in section 4.3.2. Process TP waste outputs are surrounded by a black square, the TP inputs are surrounded by a green square and the product stream by a blue square.

National annual mean discharge limits to water courses for TP can be set as low as 2 mg(P) l^{-1} and since W_3 falls below this threshold P removal may not be necessary [Carty *et al.* 1997; Hammer and Hammer 2004; EPA 2013]. It was also estimated that approximately 59 mg of TP per production run was emitted in the solid waste (SW), most likely due to P-containing compounds found within the cellular debris and residual P nutrients from the media. The final recombinant protein product contains 16.9 mg of TP which can be attributed to the P salts of the sodium phosphate buffer employed during the last purifications step.

5.4.2.2 Total phosphorus mass balance for recombinant protein production using affinity-based purification strategy

Total phosphorus analysis was carried out on waste stream and product stream samples generated during the production of the recombinant β -galactosidase employing the affinity-based purification method as outlined in section 5.3.4. Figure 5.4 shows the results of the TP mass balance determination for the laboratory-scale run. Again aqueous waste stream one (W_1) emits the highest quantity of P with 213.1 mg TP per production cycle, followed by waste streams two (W_2), four (W_4) and three (W_3) with 195.5 mg, 165.4 mg and 31.2 mg TP per run, respectively. W_1 which is the spent fermentation media obtains its high TP value due to the unused P-containing nutrients and leaked P-containing cellular constituents. The remaining waste streams, produced during the product isolation steps, include buffering components with P salts which gives these wastes their high P output. All aqueous waste streams produced during this process exhibit TP concentrations greater than typical urban wastewater and will therefore require pretreatment prior to discharge to the local WWTP [Carty *et al.* 1997]. The solid waste (SW) is estimated to contain approximately 59 mg of TP, which again is a result of the cellular debris with incorporated P as well as remaining P-containing media nutrients and the P salts from the buffer used during cell lysis. The TP in the β -galactosidase product solution is 13.4 mg which can also be attributed to the P salts of the sodium phosphate buffer employed during SEC.

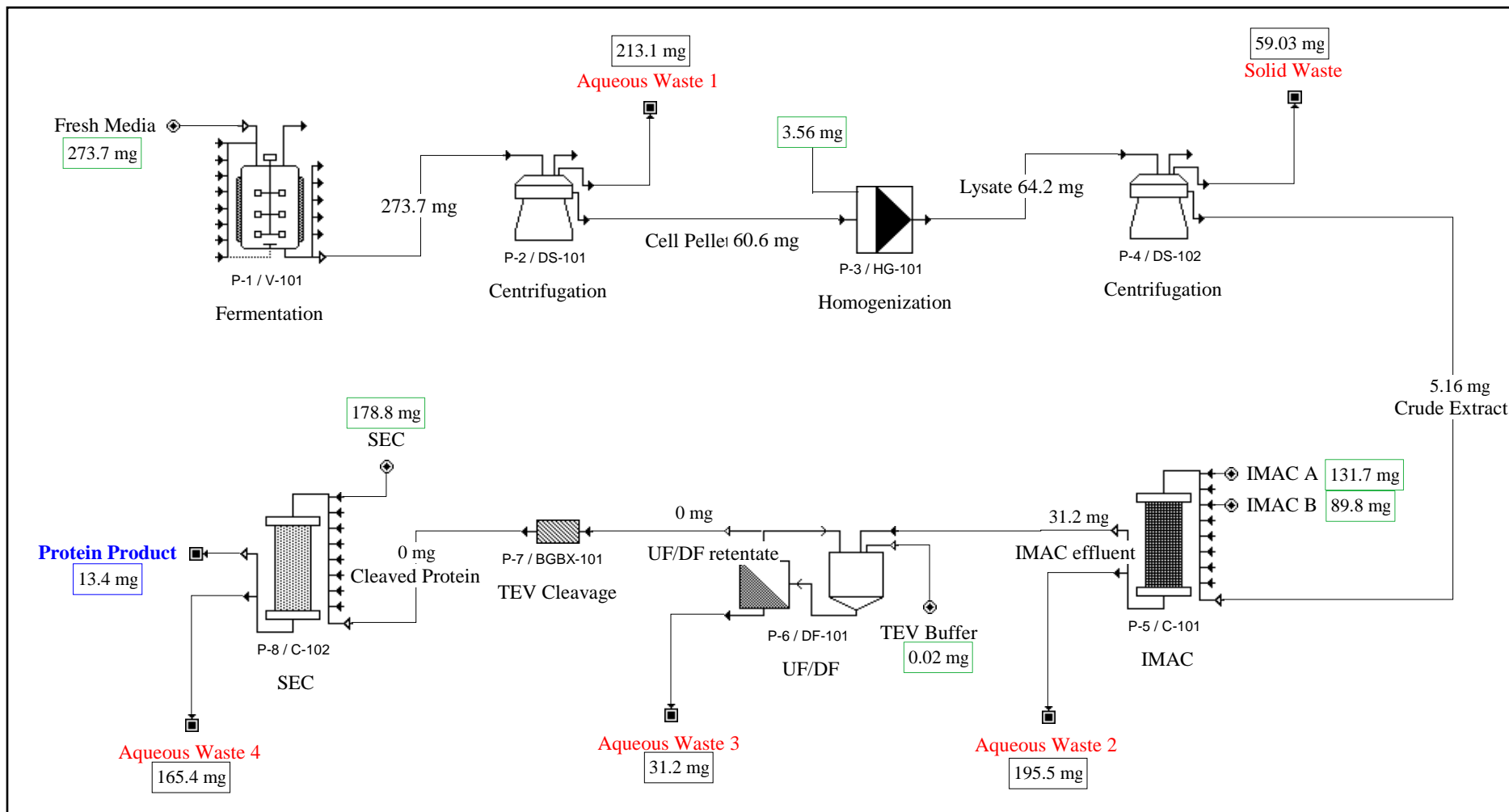


Figure 5.4 Flowsheet detailing TP mass balance for the recombinant protein production employing the affinity-based purification strategy

Results present TP in mg TP waste volume per production run at laboratory-scale as outlined in section 4.3.2. Process TP waste outputs are surrounded by a black square, the TP inputs are surrounded by a green square and the product stream by a blue square.

5.4.2.3 Total phosphorus mass balance for recombinant protein production using P-reduced fermentation media and the affinity-based purification scheme with the alternative buffering system

Waste stream and product stream samples collected during the laboratory-scale production of the recombinant model strain using the 70 % P-reduced TB media in conjunction with the affinity-based purification method with the alternative buffering system were analysed for TP as outlined in section 5.3.4. The findings of the TP mass balance analysis are illustrated in Figure 5.5. The liquid waste stream with the highest P concentration is still the spent fermentation media (W_1) with an overall TP output of 87.9 mg per laboratory-scale production run, followed by waste streams two (W_2), four (W_4) and three (W_3) with 1.44 mg, 0.029 mg and 0.026 mg TP per manufacturing cycle, respectively. The P-optimised media resulted in a 58.6 % reduction in TP observed in aqueous waste stream one (W_1) indicating an improvement from an environmental viewpoint. The majority of remaining TP emitted during this manufacturing process was in the solid waste (SW) with 22.95 mg TP per production. The P content in the SW is a result of P-containing membrane lipids and some residual P-comprising nutrients from the spent media. The other waste streams (W_2 - W_4) are unlikely to require any pretreatment for P removal prior to their release to the local WWTP depending on the facilities IPPC license agreement. The P concentration in W_3 and W_4 is so low (less than $0.3 \text{ mg(TP) l}^{-1}$) that a further reduction in P is not necessary before discharge into a water course [Carty *et al.* 1997]. As the final step during this protein purification process consists of SEC using the ACES buffer, there is a negligible amount of TP (0.003 mg(TP) per run) in the product due only to the added trypsin inhibitor.

These findings demonstrate that this process, i.e. the P-reduced fermentation media in combination with the alternative buffering system used during the affinity-based purification, results in greatly reduced concentrations of TP in the waste streams (compared to Figure 5.3 and 5.4) and is more favourable in terms of P emissions to the environment [Eaton and Franson 1995]. Specific P removal/recovery at the manufacturing plant may only be required for aqueous waste stream one (W_1) depending on the plants emission limit values as set by the IPPC license agreement [EPA 2013].

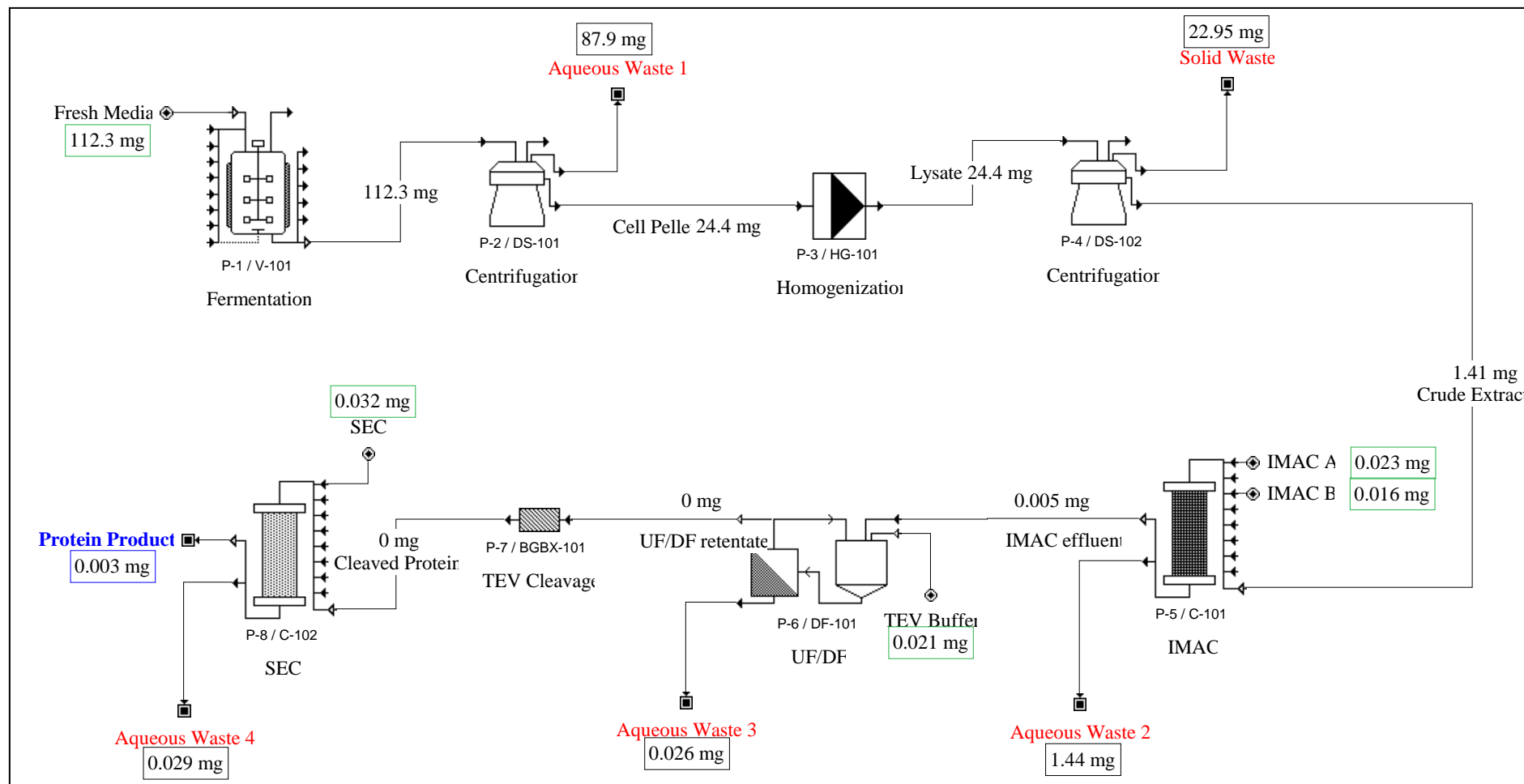


Figure 5.5 Flowsheet detailing TP mass balance for the recombinant protein production employing P-reduced TB media in combination with the affinity-based purification strategy and the alternative buffering system

Results present TP in mg TP waste volume per production run at laboratory-scale as outlined in section 4.3.2. Process TP waste outputs are surrounded by a black square, the TP inputs are surrounded by a green square and the product stream by a blue square.

Discharge limits for wastes emitted containing TP are specific for each individual industrial facility and depend upon the local WWTP or the sensitivity of the receiving waterbody [Carty *et al.* 1997; EPA 2008]. Reducing the concentration of TP in the waste streams as well as decreasing the number of wastes that require P removal prior to release would reduce the raw material and waste treatment costs in addition to benefiting the environment in terms of savings in raw material usage and averting discharge of large quantities of phosphorus as a result of an accidental facility releases or sub-optimal operating wastewater treatment system.

5.4.2.4 Total phosphorus in the waste streams of all production processes

Table 5.4 presents the combined total phosphorus results for all three recombinant protein production methods. The findings highlight the TP savings in the aqueous waste streams that can be achieved when reducing the P concentration in the fermentation media and replacing the sodium phosphate salts with ACES during downstream processing. As outlined previously, this highest amount of phosphorus in all three processes is discharged within aqueous waste stream one (W_1) which is due to the remaining P-compounds in the spent fermentation media. However, reducing the P concentration in the media led to a significant drop of TP in W_1 , which decreases the environmental burden of this waste stream. The TP concentration in the remaining liquid wastes depends on the buffering components employed for the specific unit operation, i.e. if phosphate salts were used then the TP concentration is high, as in the case of HIC, SEC and IMAC in the traditional multi-chromatographic and affinity-based purification scheme. Consequently, once these salts were replaced by ACES, the TP concentrations in the resulting waste streams were considerably reduced. Overall, the process with the greatest TP amounts discharged is the production method employing the affinity-based purification scheme with a mean TP value of 602.87 mg(TP) per run. This is followed by the production procedure using the traditional multi-chromatographic purification strategy with a total mean TP value of 555.29 mg(TP) per run. The P-reduced production process, i.e. employing the 70 % P-reduced TB media during fermentation and using the alternative buffer during the affinity-based purification, had the lowest overall TP quantity for all liquid waste streams combined with 89.9 mg(TP) per production run. This is an 85 % reduction of generated TP wastes compared the production process with the affinity-based purification strategy.

Table 5.4 Total phosphorus (TP) analyses of the aqueous waste streams generated from all three laboratory-scale recombinant protein production processes

Aqueous waste streams	TP per process volume emitted (in mg of TP per production run)		
	Traditional multi-chromatographic purification scheme	Affinity-based purification scheme	P-reduced fermentation media and affinity-based scheme with alternative buffer
W ₁	213.15 ± 9.55 ^a	213.15 ± 9.55 ^a	87.90 ± 1.23 ^a
W ₂	2.28 ± 0.07 ^b	195.11 ± 2.86 ^g	1.92 ± 0.04 ^g
W ₃	0.03 ± 0.0007 ^c	31.25 ± 0.34 ^h	0.05 ± 0.002 ^h
W ₄	150.42 ± 2.57 ^d	163.36 ± 3.11 ⁱ	0.03 ± 0.004 ⁱ
W ₅	25.59 ± 0.41 ^e		
W ₆	163.82 ± 2.00 ^f		
Total mean TP value (in mg(TP) per run)	555.29	602.87	89.90

The data shows mean TP value ± standard deviation (n = 6)

^aper 98 ml emitted; ^bper 271 ml emitted; ^cper 149 ml emitted; ^dper 129 ml emitted; ^eper 22 ml emitted; ^fper 138 ml emitted; ^gper 164 ml emitted; ^hper 121 ml emitted; ⁱper 139 ml emitted

W₁: aqueous waste stream one; W₂: aqueous waste stream two; W₃: aqueous waste stream three; W₄: aqueous waste stream four; W₅: aqueous waste stream five; W₆: aqueous waste stream six

This can significantly reduce the environmental impact of this specific production process in terms of phosphorus emissions.

5.4.3 Chemical oxygen demand of waste streams generated during recombinant protein production

The chemical oxygen demand (COD) analysis was carried out on all aqueous waste streams for each of the recombinant β -galactosidase production processes as outlined in section 5.3.4 and the results are shown in Table 5.5. COD is a measure of the amount of oxygen required to oxidise the organic matter content in a waste sample and therefore indicates its organic strength and pollution potential in natural waters [Eaton and Franson 1995; Hammer and Hammer 2004]. With the exception of aqueous waste stream five (W_5) for the process using the traditional multi-chromatographic purification, all waste streams are considered to contain medium to high strength COD concentrations and will require treatment to reduced the values to $\leq 125 \text{ mg(O}_2\text{) l}^{-1}$, which is the national threshold limit for COD emissions into the environment [Carty *et al.* 1997; Tchobanoglous *et al.* 2003].

The highest COD values for all three processes were observed in aqueous waste stream one (W_1). This can be attributed to unused nutrients from the media employed during fermentation as well as the biological components present which include proteins, cellular debris and residual suspended cells. There is no significant effect on the COD values upon reducing the P salts in the media as the organic compounds were not altered. In addition, the 70 % P-reduced media produced the same amount of biomass (section 3.4.3.2) and therefore an equal quantity of biological components would be found in each of the waste streams. The COD concentrations in the second liquid waste streams (W_2) are also relatively similar with W_2 from the process employing the traditional multi-chromatographic purification scheme being the highest with 3.43 g of O_2 per production run, followed by the waste streams containing the ACES buffer with 3.30 g of O_2 per run and then W_2 from the process using the affinity-based method employing the phosphate buffer with 2.69 g of O_2 per run. The ACES salt used during the IMAC purification step influences the oxygen balance in the waste stream, thereby, increasing the COD value by 23 % compared to W_2 produced during the affinity-based process which mainly contains P salts.

Table 5.5 Chemical oxygen demand (COD) analyses of the aqueous waste streams generated from all three laboratory-scale recombinant protein production processes

Aqueous waste streams	COD per process volume emitted (in g of O ₂ per production run)		
	Traditional multi-chromatographic purification scheme	Affinity-based purification scheme	P-reduced fermentation media and affinity-based scheme with alternative buffer
W ₁	6.95 ± 0.33 ^a	6.95 ± 0.33 ^a	6.43 ± 0.25 ^a
W ₂	3.43 ± 0.19 ^b	2.69 ± 0.14 ^g	3.30 ± 0.11 ^g
W ₃	1.36 ± 0.04 ^c	1.35 ± 0.02 ^h	1.47 ± 0.02 ^h
W ₄	0.05 ± 0.02 ^d	0.08 ± 0.004 ⁱ	0.58 ± 0.02 ⁱ
W ₅	0.001 ± 0.0002 ^e		
W ₆	0.05 ± 0.001 ^f		
Total mean COD value (in g(O₂) per run)	11.84	11.07	11.78

The data shows mean COD value ± standard deviation (n = 6)

^aper 98 ml emitted; ^bper 271 ml emitted; ^cper 149 ml emitted; ^dper 129 ml emitted; ^eper 22 ml emitted; ^fper 138 ml emitted; ^gper 164 ml emitted; ^hper 121 ml emitted; ⁱper 139 ml emitted

W₁: aqueous waste stream one; W₂: aqueous waste stream two; W₃: aqueous waste stream three; W₄: aqueous waste stream four; W₅: aqueous waste stream five; W₆: aqueous waste stream six

In all three processes, the aqueous waste streams generated during the UF/DF step are also fairly similar with values of 1.36 g of O₂ per run, 1.35 g of O₂ per run and 1.47 g of O₂ per run for W₃ using the traditional multi-chromatographic strategy, affinity-based purification scheme and the P-reduced method, respectively. The COD values in these waste streams are mainly a result of the buffering components employed and to a smaller extent the trace amount of the biological compounds. In both affinity-based purification methods, imidazole was used during the IMAC purification step, resulting in a high N content in the waste streams. However, these findings show that the N compounds were readily degraded. The remaining waste streams (W₄-W₆) that arise from the traditional multi-chromatographic purification scheme as well as W₄ from the affinity-based production process contain low concentrations of COD (between 0.001-0.08 g of O₂ per laboratory-scale production run). This is because the unit steps producing these waste streams used sodium phosphate-based buffering systems, resulting in low COD values. The COD that could be detected was due to carryover and the biological waste constituents as well as the trypsin inhibitor present in all buffers. All of these waste streams fall within the range of typical national urban wastewater COD values (0.25-0.80 g(O₂) l⁻¹) and therefore, depending on the specific ELV in the facilities IPPC license, are unlikely to require pretreatment prior to discharge to the local WWTP [Carty *et al.* 1997]. In the last waste stream (W₄) obtained during the production process with the affinity-based purification scheme employing the alternative buffering system, the COD value is a little higher compared to W₄ for the affinity-based process (0.58 versus 0.08 g of O₂ per run). This is as a result of the ACES salt used (in place of sodium phosphate) in the buffer for the SEC step.

This analysis indicates that the majority of COD exits the process in the first three waste streams for all three simulations. The findings also show that replacing the sodium phosphate buffer with ACES increases the overall COD output by only 6 % for the total production process which is still lower compared to the production method employing the traditional multi-chromatographic purification scheme. Generally, the total mean COD values for all three processes are relatively similar ranging from 11.07 to 11.87 g of O₂ per run.

5.5 Conclusions

In this chapter the full recombinant protein production process using the model β -galactosidase protein designed was evaluated and the wastes that arise during manufacturing employing the three different processes were compared.

Overall, the findings show that for all processes investigated the waste stream with the highest pollution load is aqueous waste stream one (W_1), i.e. the spent media. This can be attributed to residual unused nutrients and biological components present after fermentation. These waste streams contain both high COD as well as TP values identifying them as environmental “hot spots” which may require monitoring and treatment. The results also indicated that employing a 70 % P-reduced fermentation media can significantly reduce the phosphorus outputs in that particular waste stream. In addition, the mass balance analysis demonstrated that replacing the sodium phosphate salts used in the buffers during downstream processing with an ACES buffering system greatly reduces the P outputs in the waste streams generated during the purification steps. In combination with the P-reduced fermentation media, an 85 % reduction of overall TP wastes could be achieved when employing the production method with the affinity-based purification system, whereas, little impact on the overall COD levels of the waste streams was observed.

Phosphorus reductions in the upstream (fermentation) and downstream processing (purification) steps during the manufacturing of recombinant proteins will evidently result in reduced TP quantities in the waste streams generated. Consequently, these P reductions will affect the costs involved in waste treatment, either due to eliminating P removal prior to discharge of the wastewater or reducing the amount of P-containing sludge that arises during P removal. Furthermore, the environmental impact, if accidental discharge at the facility was to occur, is significantly reduced.

**Chapter Six – Environmental modelling of the
phosphorus emissions from recombinant protein
production processes using the EHS model**

6.1 Introduction

Traditionally, the development of a biopharmaceutical production strategy focussed on operational efficiency which subsequently impacts upon costs, time and economic profits. In recent years, escalating environmental control costs and newly issued environmental regulations, resulted in increasing interests in improving the environmental performance of the processes. More attention is being paid to the minimisation of pollution through implementation of cleaner technologies or processes as it reduces the need for end-of-pipe technologies which are generally used to lower the amount of harmful emissions and substances. Early process development aims now include the best ecological process for further development which facilitates the integration of environmental perspectives into industrial process design via simulation models incorporating environmental impact assessments [Farid 2007; Qian *et al.* 2007; Li *et al.* 2009].

The previous chapters in this thesis have highlighted that significant quantities of pollutants, in particular phosphorus (P), are emitted in the wastes generated during the production of a recombinant protein. Life cycle assessments (LCA) could be used for a detailed characterisation of the environmental impacts of these wastes arising from the recombinant β -galactosidase manufacturing processes as described in this study [Li *et al.* 2009]. Additionally, a simplified LCA approach will be able to evaluate and compare the modifications made to the process, i.e. P-reduced fermentation media, purification set-up and buffering system, and determine the most environmentally friendly process design.

Life cycle assessments can be achieved by utilising an environmental modelling technique. A number of different models have been developed to deal with various broad and/or specific environmental decisions (e.g. [Seppälä *et al.* 1998; Zhang *et al.* 2006; Shaked and Jolliet 2011; Yan *et al.* 2011; Dandres *et al.* 2012; Mayyas *et al.* 2012; Marvuglia *et al.* 2013]). These include models which specifically analyse the environmental burden of industrial activities, as well as assess techniques which reduce the impact or optimise the process (e.g. [Khan *et al.* 2001; Ramzan *et al.* 2008; Pietrzykowski *et al.* 2013; Wang *et al.* 2013]).

The ecological hazard score (EHS) model by Cliffe [2010] was employed in this study to evaluate the environmental impacts of the waste streams generated by a recombinant protein production facility. This will show the environmental burden of the whole manufacturing process as well as establish specific waste streams or unit operations for environmental improvement opportunities. In addition, employing the EHS model will demonstrate that environmental life cycle assessments can be utilised for process optimisations rather than to a product as it has traditionally been used [Li and Hui 2001; Qian *et al.* 2007].

The environmental burden of a manufacturing process can be determined in terms of the environmental impacts of the waste components [Hui *et al.* 2003]. The EHS model was based on the Environmental Impact Evaluation (EIE) model from Li and Hui [2001] which evaluates the environmental impact of general manufacturing processes via analysing the waste constituents. The structure of the EHS model consists of five main steps which are shown in Figure 6.1 and are described in detail in the following section (6.2).

The EHS model was applied to evaluate the environmental performance of the three laboratory-scale recombinant β -galactosidase production processes as described in previous chapters:

1. Fermentation of model *E. coli* strain with TB media and downstream processing using the traditional multi-chromatographic purification scheme (illustrated in Figure 5.1 in section 5.4.1.1)
2. Fermentation of model *E. coli* with TB media and downstream processing employing the affinity-based purification strategy (presented in Figure 5.2 in section 5.4.1.2)
3. Fermentation of model *E. coli* with 70 % P-reduced TB media and downstream processing via affinity-based purification using the alternative buffering system (P-reduced production process; outlined section 5.4.1.3)

This will provide information on the overall environmental effects of each process and identify potential opportunities for environmental improvements.

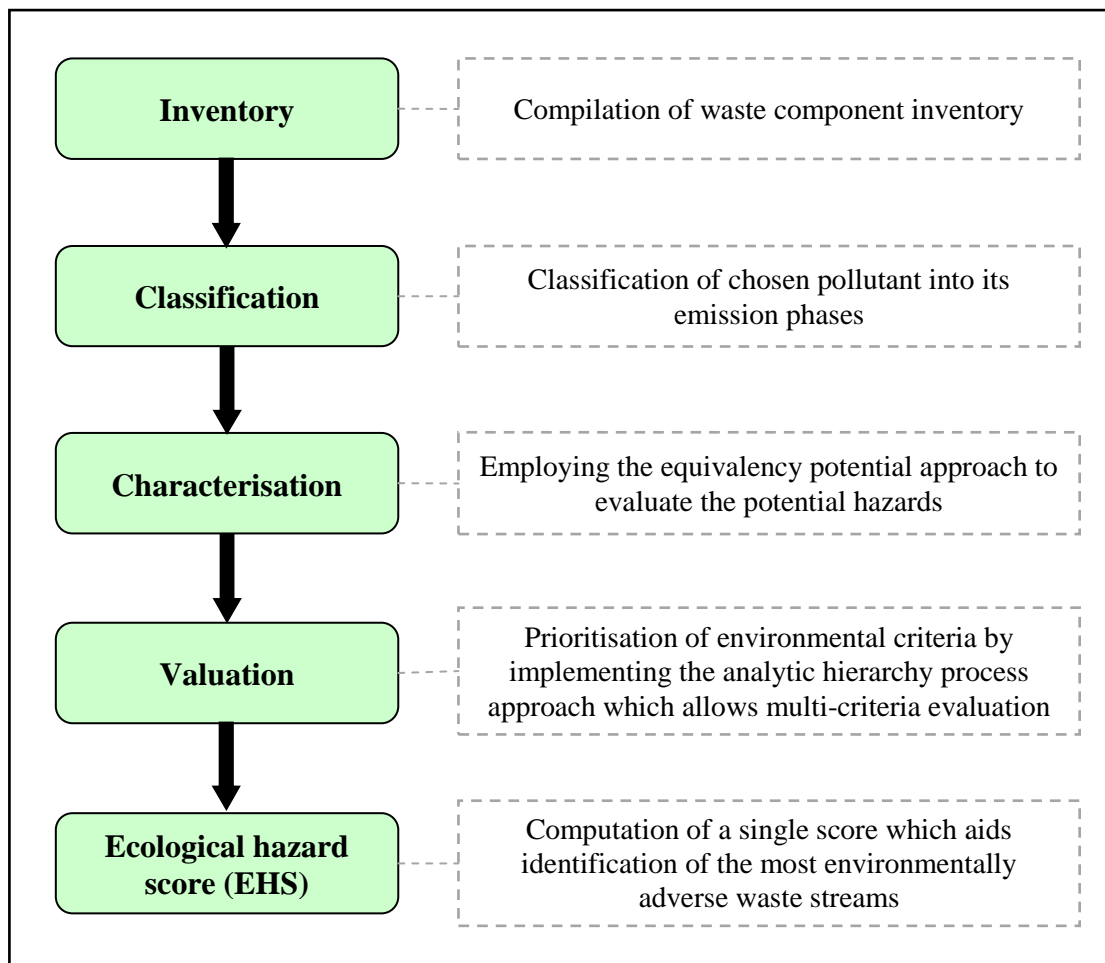


Figure 6.1 Flowsheet depicting the steps involved in the ecological hazard score (EHS) model

6.2 Ecological hazard score (EHS) model structure

The ecological hazard score (EHS) model allows quantification of the environmental impacts of a unit operation through comprehensive analysis of the waste components of the process. The methodology developed for this model consists of three main phases (section 1.7): goal and scope definition, inventory analysis and impact assessment with interpretation of the findings at each stage [Bhander *et al.* 2003; SAIC 2006]. The individual steps in the EHS model are inventory analysis, classification, characterisation, valuation and determination of a final ecological hazard score. Figure 6.2 outlines how these steps are incorporated within each of these phases. A revised computer simulation of the EHS model using Microsoft® Office Excel® 2007, originally constructed by Cliffe [2010], was employed as it maximises clarity at each step. The full system procedure for incorporation of these steps is described in the following sections.

6.2.1 Goal and scope definition

The first step in all life cycle assessments is the goal and scope definition phase as it identifies and describes the purpose of the assessment. This includes a determination of the specific system (e.g. recombinant protein production in *E. coli*) as well as the system boundaries [Wenzel *et al.* 2000; SAIC 2006]. System boundaries define the beginning and end points of analysis for the system, thereby, essentially separating the system from its surroundings. The definition of system boundaries is critical for obtaining reliable results and for the interpretation of a life cycle assessment. Two LCAs of the same product/process can result in different findings depending on their system boundaries. The requirements of the LCA as well as the specific circumstance will influence the appointment of the system boundaries [Lee *et al.* 1995; Suh *et al.* 2003].

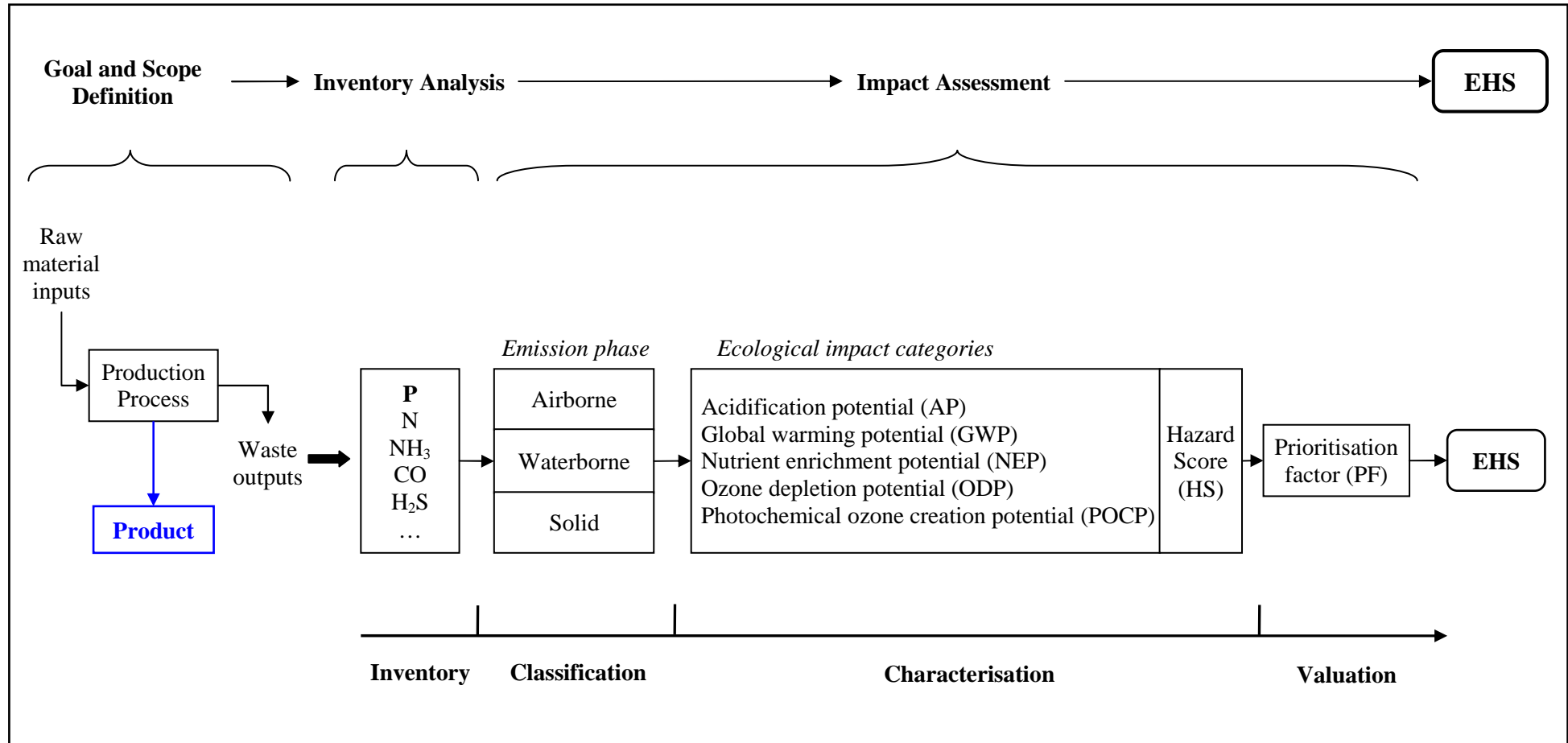


Figure 6.2 The framework of the ecological hazard score (EHS) model

Adapted from Li and Hui [2001]

6.2.2 Life cycle inventory analysis

This analysis involves the quantification of the inventory flow for a product system, i.e. the raw material inputs and emission outputs for a process [SAIC 2006]. All data necessary for construction of the model related to the functional unit defined in goal and scope definition for all activities within the system boundaries are collected [Bhander *et al.* 2003; Rebitzer *et al.* 2004]. For the EHS model a single pollutant, such as phosphorus, is selected for subsequent application in the model.

The wastes are analysed to quantify the specific pollutant under investigation and a pollutant inventory (Q) of the inputs and outputs is compiled.

6.2.3 Classification

Classification in the EHS model is the first step in the impact assessment phase and refers to the assignment of outputs to the impact categories based on scientific analysis of relevant environmental processes, i.e. classification into their emission phases [Li and Hui 2001; Bhander *et al.* 2003]. The environmental effect of individual waste constituents can often depend on the emission phase (solid, liquid or gaseous).

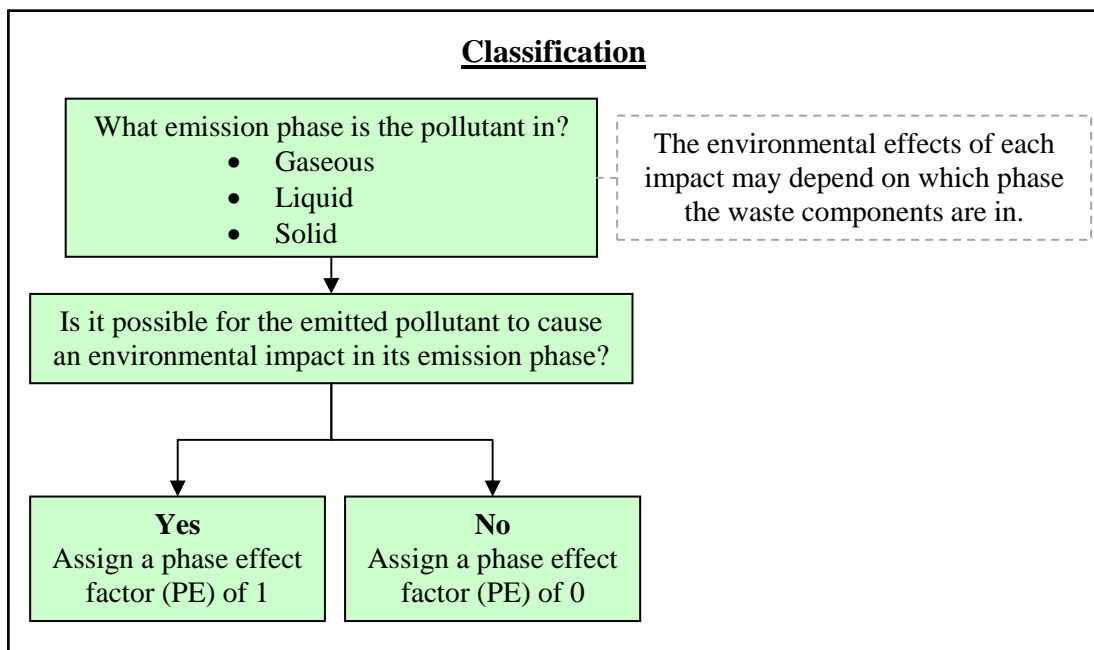


Figure 6.3 Schematic of the stages involved in the classification step for the determination of the ecological hazard score (EHS)

Waste discharged in aqueous form, for example, is likely to have a greater impact in relation to waterborne emissions than airborne emissions. Consequently, if a waste component is classified to cause an environmental impact, a phase effect factor (PE) of 1 is assigned while wastes with clearly improbable environmental impact are given a PE value of 0 [Li and Hui 2001]. Figure 6.3 visualises the classification step for the determination of the EHS.

6.2.4 Characterisation

Characterisation in the EHS model is primarily a quantitative step employing science-based conversion factors to determine the environmental impacts on ecological health. This is carried out using the equivalency potential approach during which a hazard score (HS) is calculated for each waste stream from the process evaluated in terms of the environmental impact category selected. During the first stage the scale of the contribution for each emission to the various environmental impact categories is established [SAIC 2006]. The five typical environmental impact categories considered as part of the EHS model are:

- Nutrient enrichment potential (NEP)
- Acidification potential (AP)
- Global warming potential (GWP)
- Ozone depletion potential (ODP)
- Photochemical ozone creation potential (POCP)

There are a number of environmental effect groups that can be used to evaluate environmental performance during life cycle impact assessments including ecotoxicity as well as human toxicity impact categories [Haes *et al.* 1999; Li *et al.* 2009; Lozano *et al.* 2010; Zhou *et al.* 2011]. However, the aim of this research element was directed upon the ecological effects of the waste streams and therefore human toxicity impact categories were not included in the model. If the goal and scope definition requires the additional waste production and resource consumption categories, these can be included in the assessment [Wenzel *et al.* 2000].

For each of the five categories listed above, equivalency factors (EF) are calculated by comparing the impact of the component analysed in relation to the impact of a

reference component for that category. This reference component is the dominant factor (typical pollutant) in each group. For example, in the GWP impact group the reference unit is carbon dioxide (CO₂) and consequently, the EFs express the substances' potential impact as grams of CO₂ equivalent per gram of substance (g(CO₂ eq) g(substance)⁻¹). For instance, 1 g of methane has a stronger global warming effect and contributes to it as the emission of 25 g of CO₂ would, giving it an EF of 25 [Hauschild and Wenzel 1998; Wenzel *et al.* 2000; Zhou *et al.* 2011]. The impact calculated for the compound reflects a potential impact that could arise not an actual one and the real occurrence of an impact will depend on several aspects, primarily related to the site characteristics [Mendes *et al.* 2003]. Table 6.1 presents the impact categories evaluated during this study, their associated definitions and formula for EF calculations applied. Detailed descriptions of the ecological impact categories and the complete scientific background for obtaining EFs is outlined in Hauschild and Wenzel [1998] and Wenzel *et al.* [2000].

The emissions investigated in this chapter impact mainly upon nutrient enrichment and as a result specifically the steps involved in the calculation of the nutrient enrichment potential (NEP) equivalency factor are illustrated in Figure 6.4. For a substance to contribute to the nutrient enrichment, biologically available nitrogen (N) or phosphorus (P) must be present. Therefore, the EF for NEP is determined by assessing the amount of N or P released into the environment from the wastes analysed. The calculation of the equivalency factor for any nutrient enriching substance's P/N potential thus expresses its respective contents by weight of either one of the elements [Hauschild and Wenzel 1998; Wenzel *et al.* 2000]. Figure 6.4 describes the computation of the equivalency factor for phosphate (PO₄³⁻) using the formula from Table 6.1 for P (EF(P)) and showing a result of 0.33 gram of P equivalents per gram of waste emitted.

Once the waste mass constituent inventory is assembled (section 6.2.2), the waste components are classified into their emission phases (section 6.2.3) and the equivalency factor for the pollutant under investigation is calculated, the hazard score (HS) for each of the emissions selected from the process can be determined as outlined in Figure 6.4 [Li and Hui 2001].

Table 6.1 The category quantitative assessment of the ecological impact indices

Environmental impact category	Definition	Formula
Nutrient enrichment	An equivalent quantity of reference substance (total P or total N)	$EF (P) = \frac{\pi * AM(P)}{MM (X)}; EF (N) = \frac{v * AM(N)}{MM(X)}$
Acidification	AP is expressed as an equivalent quantity of a reference substance (SO ₂)	$EF (AP) = \frac{n}{2 * MM(X)} * MM(SO_2)$
Global warming	The potential contribution to global warming from a given quantity of the gas relative to the contribution for a corresponding quantity of carbon dioxide (CO ₂)	$GWP_1 = \frac{\text{contribution to GW from gas (i) over T years}}{\text{contribution from CO}_2 \text{ to GW over T years}}$
Ozone depletion	Individual gases express ODP as an equivalent emission of a reference substance trichlorofluoromethane (CFC-11)	$ODP_1 = \frac{\text{contribution to stratospheric ozone depletion from gas (i)}}{\text{contribution to stratospheric ozone depletion from CFC-11}}$
Photochemical ozone creation	POCP values express the ozone formation potential as an equivalent emission of a chosen reference substance ethylene (C ₂ H ₄)	$POCP_1 = \frac{\text{contribution to ozone formation from gas (i)}}{\text{contribution to ozone formation from C}_2\text{H}_4}$

Adapted from Wenzel et al. [2000] and Li et al. [Prescott et al. 2002; 2009; Madigan et al. 2009]

: represents the number of phosphorus atoms in the substance's notational formula; : represents the number of nitrogen atoms in the substance's notational formula; AM: atomic mass; MM: molecular mass; X: pollutant under investigation; P: phosphorus; N: nitrogen; n: represents number of hydrogen ions released in the recipient as a result of conversion of the pollutant; GW: global warming; gas (i): pollutant under investigation; T: Time interval

Characterisation

Equivalency factor (EF) calculation for phosphorus

$$EF(P) = \frac{\pi * AM(P)}{MM(X)}$$

AM(P): atomic mass of phosphorus
 MM(X): molecular mass of substance in question
 π: represents number of P atoms in the substance

The equivalency factor (EF) for a nutrient enriching substance is calculated.

Table A: Equivalency factor for nutrient enrichment [Hauschild and Wenzel 1998; Wenzel *et al.* 2000]

Substance	Formula	Molecular Mass (MM) g mol ⁻¹			EF(N)	EF(P)
					g(N) g(subs.) ⁻¹	g(P) g(subs.) ⁻¹
N-compounds						
Nitrate	NO ₃ ⁻	62.00	1	0	0.23	0
Nitrogen dioxide	NO ₂	46.01	1	0	0.30	0
Nitrite	NO ₂ ⁻	46.01	1	0	0.30	0
Nitrogen oxide	NO _x ^a	46.01	1	0	0.30	0
Nitric oxide	NO	30.01	1	0	0.47	0
Ammonia	NH ₃	17.03	1	0	0.82	0
Cyanide	CN ⁻	26.02	1	0	0.54	0
Total N	N	14.01	1	0	1	0
P-compounds						
Phosphate	PO ₄ ³⁻	94.97	0	1	0	0.33
Pyrophosphate	P ₂ O ₇ ²⁻	173.94	0	2	0	0.35
Total-P	P	30.97	0	1	0	1

π: represents the number of N atoms in the substance's notational formula

π: represents the number of P atoms in the substance's notational formula

^a In NO_x, "x" is assumed to have an average value of 2

Table A shows a full list of NEP equivalency factors for N and P.

Using the information detailed in Table A, the EF for PO₄³⁻ can be computed

Example:

Nutrient enrichment equivalency factor calculation for phosphate (PO₄)

$$AM(P) = 30.97 \text{ g mol}^{-1} \quad MM(PO_4^{3-}) = 94.97 \text{ g mol}^{-1} \quad = 1 \text{ P atom}$$

$$EF(P) = \frac{1 * 30.97 \text{ g mol}^{-1}}{94.97 \text{ g mol}^{-1}} = 0.33 \text{ g(P-eq) g(PO}_4^{3-})^{-1}$$

The hazard score (HS) can be computed using equation below

$$HS = \sum (EF * Q * PE)$$

EF: equivalency factor

Q: quantity of waste emitted (section 6.2.2)

PE: phase effect factor (Figure 6.3)

Figure 6.4 Schematic overview of the stages involved in the characterisation step of the ecological hazard score (EHS) model

6.2.5 Valuation

The different ecological impact groups can have various levels of environmental effect and are therefore assigned different relative weightings [Li and Hui 2001]. This is carried out during the valuation step of the EHS model (Figure 6.2) and involves the assignment of weights to each impact category based on its perceived importance in relation to the goal of the study [SAIC 2006]. Once the different impact categories have their individual weights assigned, a single value, termed prioritisation factor (PF) in the EHS model, is summated using the weighting method [Hertwich *et al.* 1997]. In the EHS model the analytical hierarchy process (AHP) was employed as the weighting method because it is an effective approach in dealing with assessments by reducing decisions to pair-wise comparisons. Furthermore, the consistency of the judgements are checked during the AHP which aids in reducing bias [Saaty and Vargas 2012].

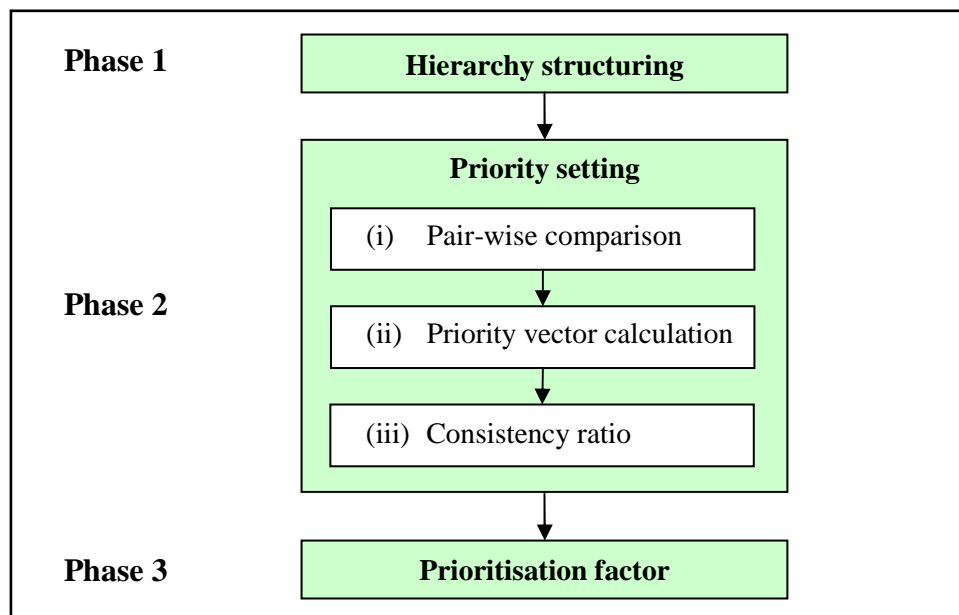


Figure 6.5 Steps of the analytical hierarchy process (AHP) application

Phase 1: Structuring of the problem as a three level hierarchy; Phase 2: (i) Pair-wise comparison is carried out between the criteria at each level of the hierarchy with respect to the preceding level, (ii) Calculation of priority vectors, (iii) Computation of the consistency ratio to ensure consistent judgements at each level; Phase 3: Determination of the prioritisation factor by multiplying the priority vector of the alternatives and the criteria

The prioritisation factor (also known as weighting factor) is determined by combining expert opinions as part of the AHP methodology. The findings can then

be used for an equal environmental evaluation and comparisons of waste streams from two or more processes. AHP is a multi-criteria analysis technique based on three principles: construction of a hierarchy, priority setting and establishing of the prioritisation factor [Ramzan *et al.* 2008]. Figure 6.5 outlines the three basic stages of the AHP application.

The first phase in the AHP approach involves structuring the problem into three levels starting with the overall objective or the goal (Level 1), followed by criteria (Level 2) and alternatives at the bottom (Level 3) as illustrated by valuation phase 1 in Figure 6.6. This structure will permit a judgement of the importance of the elements in a given level with respect to elements in the immediate level above [Yu *et al.* 2007; Ramzan *et al.* 2008; Aguilar-Lasserre *et al.* 2009; Saaty and Vargas 2012].

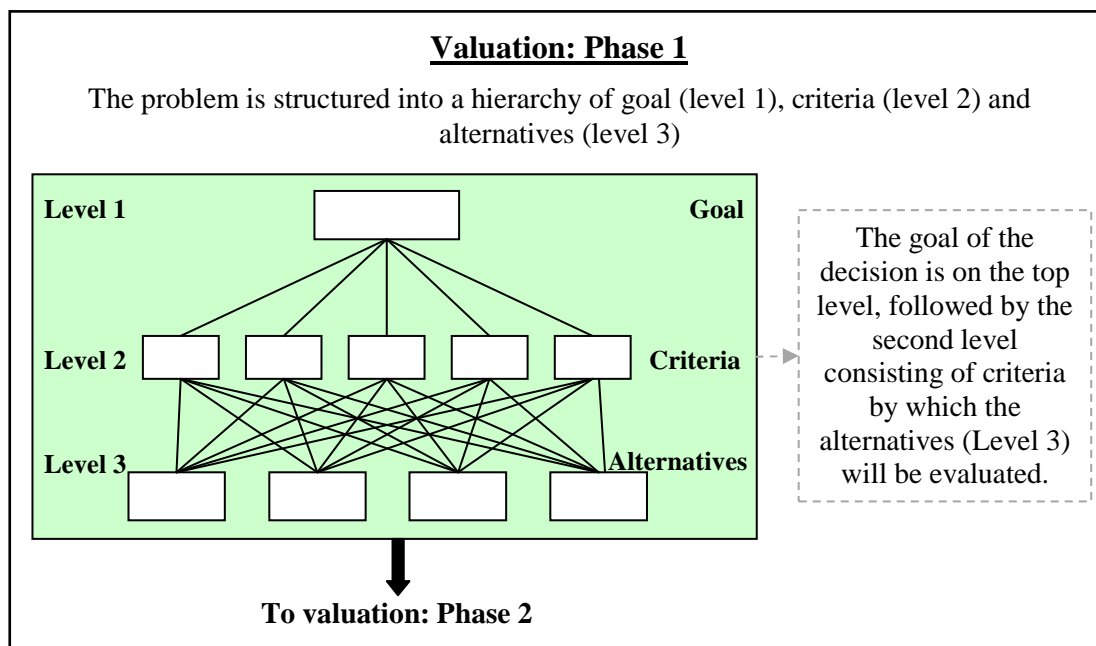


Figure 6.6 Schematic outline of phase 1 of the valuation step for the determination of the ecological hazard score (EHS)

During the next step (phase 2, part (i) Figure 6.7), the relative importance of the criteria (Level 2) are determined by performing a pair-wise comparison of criteria with respect to the goal (Level 1). Every combination of elements from a specific level is separately compared to a relevant element from a preceding level in the hierarchy.

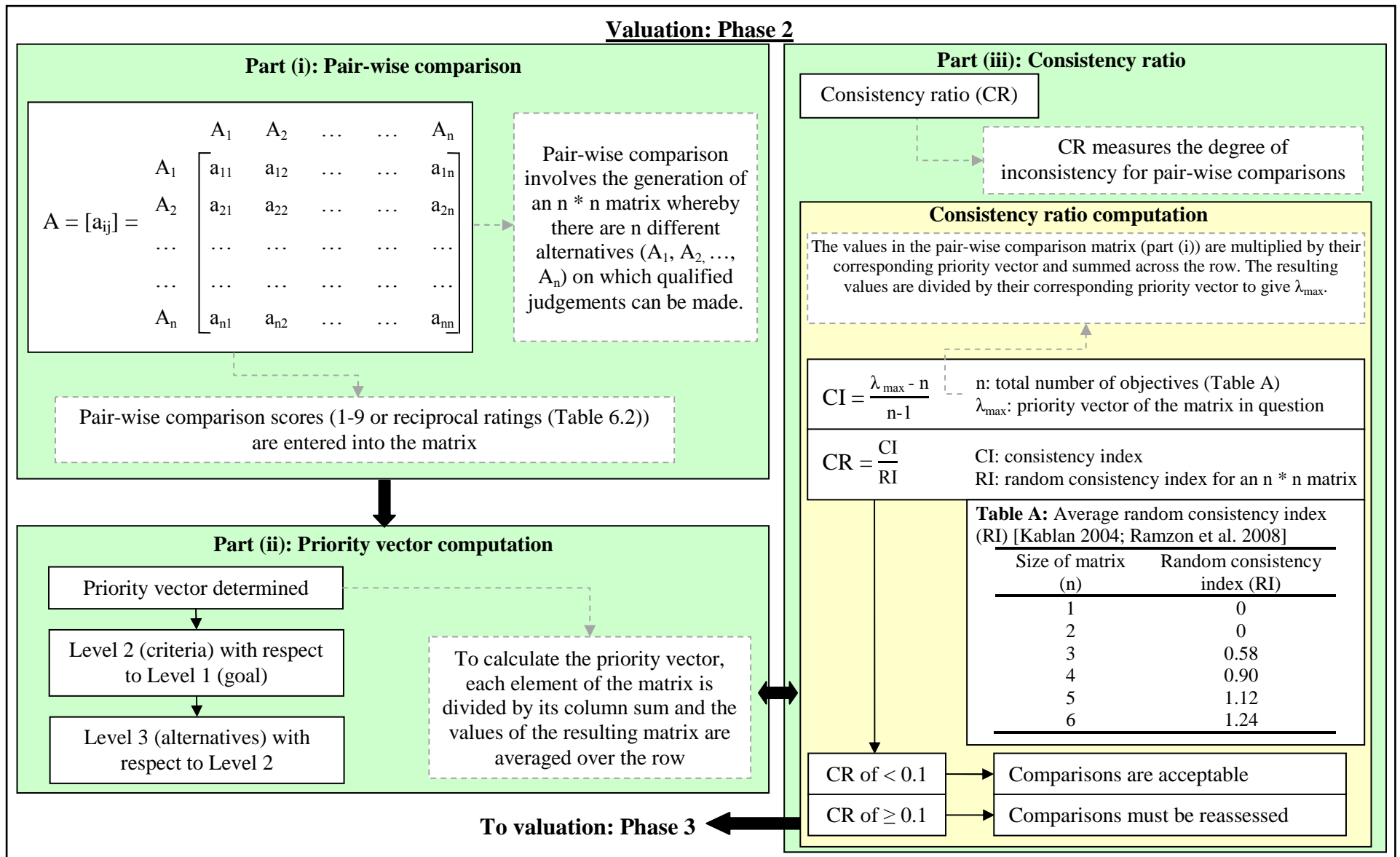


Figure 6.7 Schematic overview of phase 2 of the valuation step for the determination of the ecological hazard score (EHS)

The judgements of these pair-wise comparisons are presented in an n by n ($n * n$) matrix whereby n different and independent alternatives (A_1, A_2, \dots, A_n) are analysed. In part (i) (Figure 6.7), A is the AHP matrix while the term a_{ij} is an element of the $n * n$ matrix A . For instance, comparing alternative A_1 with alternative A_2 provides a numerical judgement a_{12} , which signifies the importance of A_1 over A_2 [Kablan 2004]. A panel of three to seven experts is required for the pair-wise comparison judgements [Li and Hui 2001; Hui *et al.* 2003].

Table 6.2 The standardised numerical comparison scale for the construction of pair-wise comparison matrices

Intensity of importance	Definition	Explanation
1	Equal importance	Two criteria contribute equally to the objective
3	Moderate importance	Experience and judgement slightly favour one criteria over another
5	Strong importance	Experience and judgement strongly favour one criteria over another
7	Very strong or demonstrated importance	One criteria is favoured very strongly over another; its dominance demonstrated in practice
9	Extreme importance	The evidence favouring one criteria over another is of the highest possible order of affirmation
2, 4, 6, 8	Intermediate values	When compromise between values of 1, 3, 5, 7 and 9 is needed

Adapted from Saaty and Vargas [2012] and Ramzan et al. [2008]

A standardised evaluation scheme transforms verbal judgements into numerical quantities using a scale from 1-9 (Table 6.2) and a reciprocal rating (i.e. $1/9, 1/8, \dots$) if the second criteria is preferred to the first. When the element is compared with itself or two components have equal importance, the value of 1 is assigned. The score 9, however, indicates affirmed evidence is present to prefer one criteria over the other [Ramzan *et al.* 2008; Saaty and Vargas 2012].

Once the $n * n$ matrix is set up, priority vectors for each criteria are calculated (part (ii), Figure 6.7). Priority vectors are also reported as eigenvectors or eigenvalues,

nevertheless, in this study they will be identified as priority vectors [Ramzan *et al.* 2008; Aguilar-Lasserre *et al.* 2009; Saaty and Vargas 2012]. For the calculation of the priority vectors, each element of the $n * n$ matrix (a_{ij}) is divided by its column sum followed by averaging the rows of the resulting matrix. This produces the computation of the priority vectors for all elements in Level 2 and 3 with respect to the preceding level [Kablan 2004].

The degree of inconsistency in decision making needs to be kept low and is determined by computing a consistency ratio (CR) (part (iii), Figure 6.7). While a good CR is < 0.1 , a ratio of ≥ 0.1 indicates the judgements made are inconsistent and the pair-wise comparison matrix should be reassessed until the CR value meets acceptable standards. The steps involved in the calculation of the CR are shown in part (iii), Figure 6.7 [Ramzan *et al.* 2008; Yang *et al.* 2012].

Phase 3 in the AHP method is the determination of the prioritisation factor (PF) via computation of the priority vectors for each level of the hierarchy, as outlined in Figure 6.8, to obtain an overall priority for each alternative [Kablan 2004; Saaty and Vargas 2012].

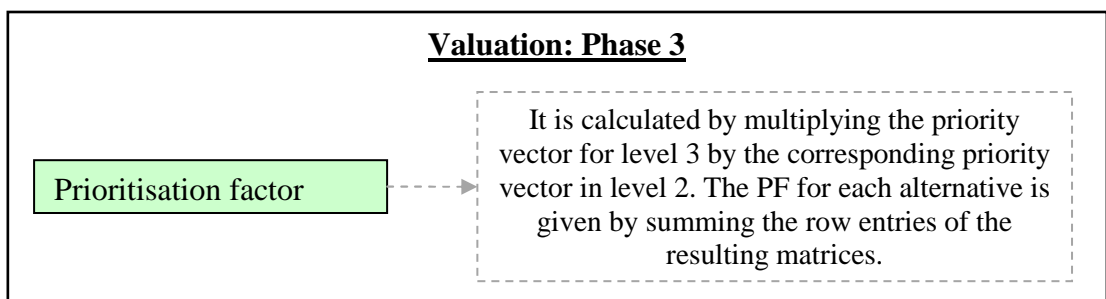


Figure 6.8 Outline of phase 3 in the valuation step for the determination of the ecological hazard score (EHS)

6.2.6 Ecological hazard score (EHS) calculation

Determination of the ecological hazard score (EHS) is the final step of the model and is achieved by multiplying the hazard score (HS) values (section 6.2.4) from each waste stream investigated by their relevant prioritisation factors (PF) (section 6.2.5) as described in Figure 6.9. Summating the EHS values obtained for each individual waste stream in a selected process will result in a single score for the entire process under investigation.

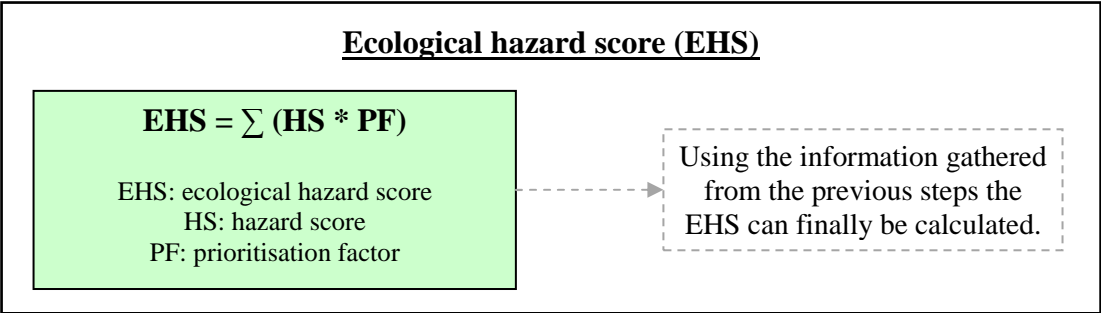


Figure 6.9 Outline of the calculation of the ecological hazard score (EHS)

The overall structure and computational methodology of the complete EHS model is summarised in Figure 6.10. The model generates an overall EHS which allows assessments on the environmental performance of a full process and its waste streams.

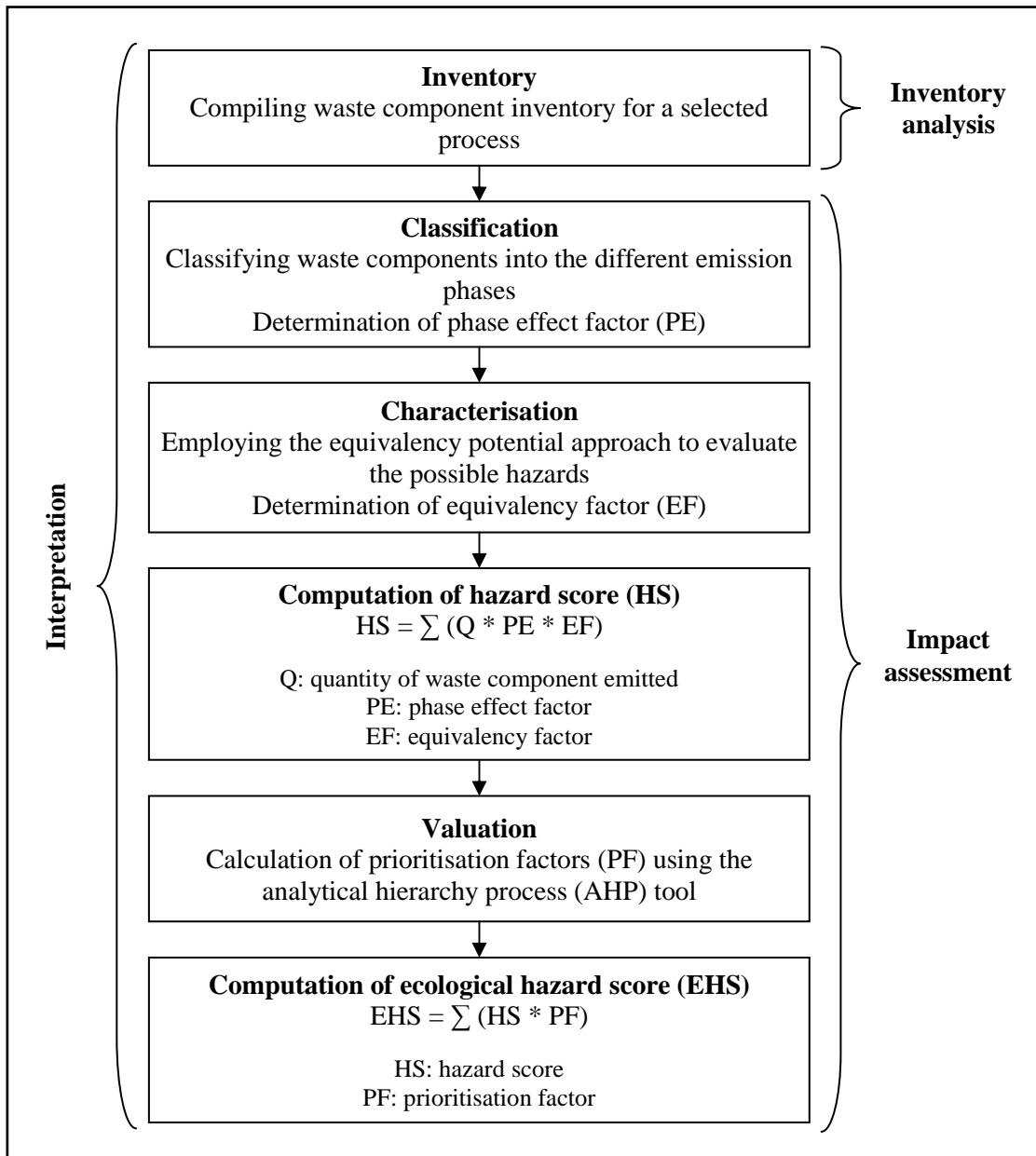


Figure 6.10 Procedure for computation of ecological hazard score (EHS)

Adapted from Li and Hui [2001]

6.3 Results and Discussions

This section shows the use of the EHS model to assess the environmental performance of each of the recombinant protein production processes designed during this study as well as their associated waste streams. The model was applied to evaluate and compare the environmental burden resulting from the β -galactosidase manufacturing processes employing the two separate purification schemes (case study one) as well as highlight the environmental saving that can be achieved by using the P-reduced production process, i.e. using 70 % P-reduced TB media and ACES salts during purification (case study two). The two case studies are visualised in Figure 6.11.

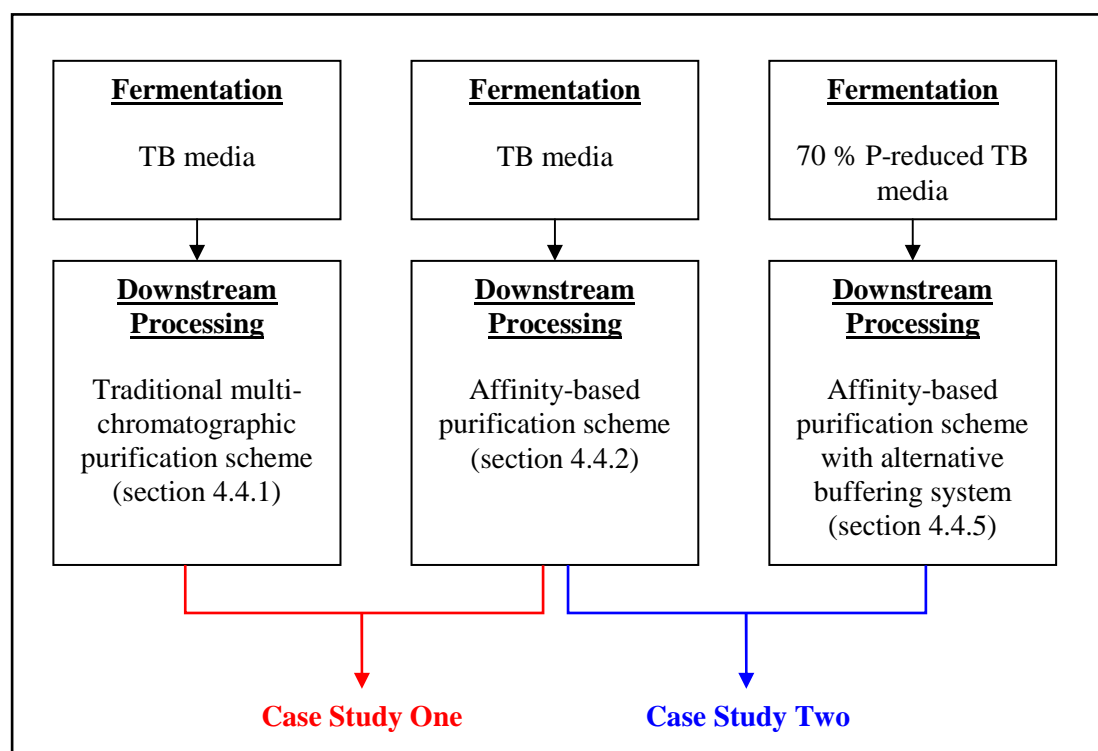


Figure 6.11 Overview showing the application of the ecological hazard score (EHS) model during this study

6.3.1 Case study one: Application of the EHS model to the recombinant protein production using traditional multi-chromatographic and affinity-based purification particularly with respect to total phosphorus (TP) emissions

The EHS model was used to assess the environmental impact of the waste streams emitted during the production of the recombinant β -galactosidase protein using the two different purification processes as outlined in chapter five of this thesis:

1. Fermentation of model *E. coli* strain with TB media and downstream processing using the traditional multi-chromatographic purification scheme (as shown in Figure 5.1 in section 5.4.1.1)
2. Fermentation of model *E. coli* with TB media and downstream processing employing the affinity-based purification strategy (as shown in Figure 5.2 in section 5.4.1.2)

Each of the processes was assessed separately using the EHS model and the individual scores allowed a comparison of the environmental performance of each production method.

6.3.1.1 Goal and scope definition

The goal of this life cycle assessment is to evaluate the potential environmental impacts from the aqueous waste streams generated during the two separate recombinant protein production processes. This includes the additional aim of determining the production process with the lowest impact on the environment particularly in terms of TP emissions. As described in chapter five, the production of recombinant proteins using the *E. coli* model strain results in a number of waste streams containing different nutrient emissions. The EHS model applied herein will facilitate an evaluation of the possible environmental impact from each waste stream in addition to comparing the total wastes emitted from the processes.

This study will investigate the environmental impacts at laboratory-scale manufacturing only. However, this should provide relevant information that can be applied during scale-up to pilot plant- and industrial-scale production of the recombinant protein. The environmental impact groups which will be used to characterise the wastes produced are outlined in section 6.2.4 and include nutrient

enrichment potential (NEP), acidification potential (AP), global warming potential (GWP), ozone depletion potential (ODP) and photochemical ozone creation potential (POCP).

The functional unit has to be clearly defined as it provides a reference to which the inputs and outputs are related, thereby, ensuring a reliable comparison between the different processes. In this case study, the waste streams from batch fermentation and purification of a recombinant model protein were analysed using the data obtained from experimental analyses (chapters three, four and five). The functional unit was defined as grams of phosphorus equivalents per gram of waste ($\text{g(P-eq) g(waste)}^{-1}$) emitted to the environment in each of the waste streams for the two processes.

The system boundary was determined according to section 6.2.1. Generally, in the biopharmaceutical industry, processes are often considered within the narrow system boundary of the process itself, also called “gate-to-gate” boundary [Van der Vorst *et al.* 2009]. This approach was applied during this study and therefore the boundaries are set to investigate the activity from the initiation of fermentation until the end of the manufacturing process. To maintain clarity, the life cycle assessment (LCA) is restricted to investigations of the core unit operations involved in recombinant protein production. The EHS model also limits the LCA to one specific pollutant within the process at a time, which in the case of this study is total phosphorus (TP). Excluded from this study are: raw material acquisition, formulation, packaging, distribution and final outcome of the biopharmaceutical protein as well as energy and fuel considerations. These restrictions were employed to maintain, as far as possible, the focus of the assessment. Figure 6.12 presents a schematic overview outlining the system boundaries, input and output points of the recombinant model protein production employing the traditional multi-chromatographic purification scheme (section 4.4.1). The system boundaries as well as input and output points of the production process using the affinity-based purification strategy (section 4.4.2) are illustrated in Figure 6.13.

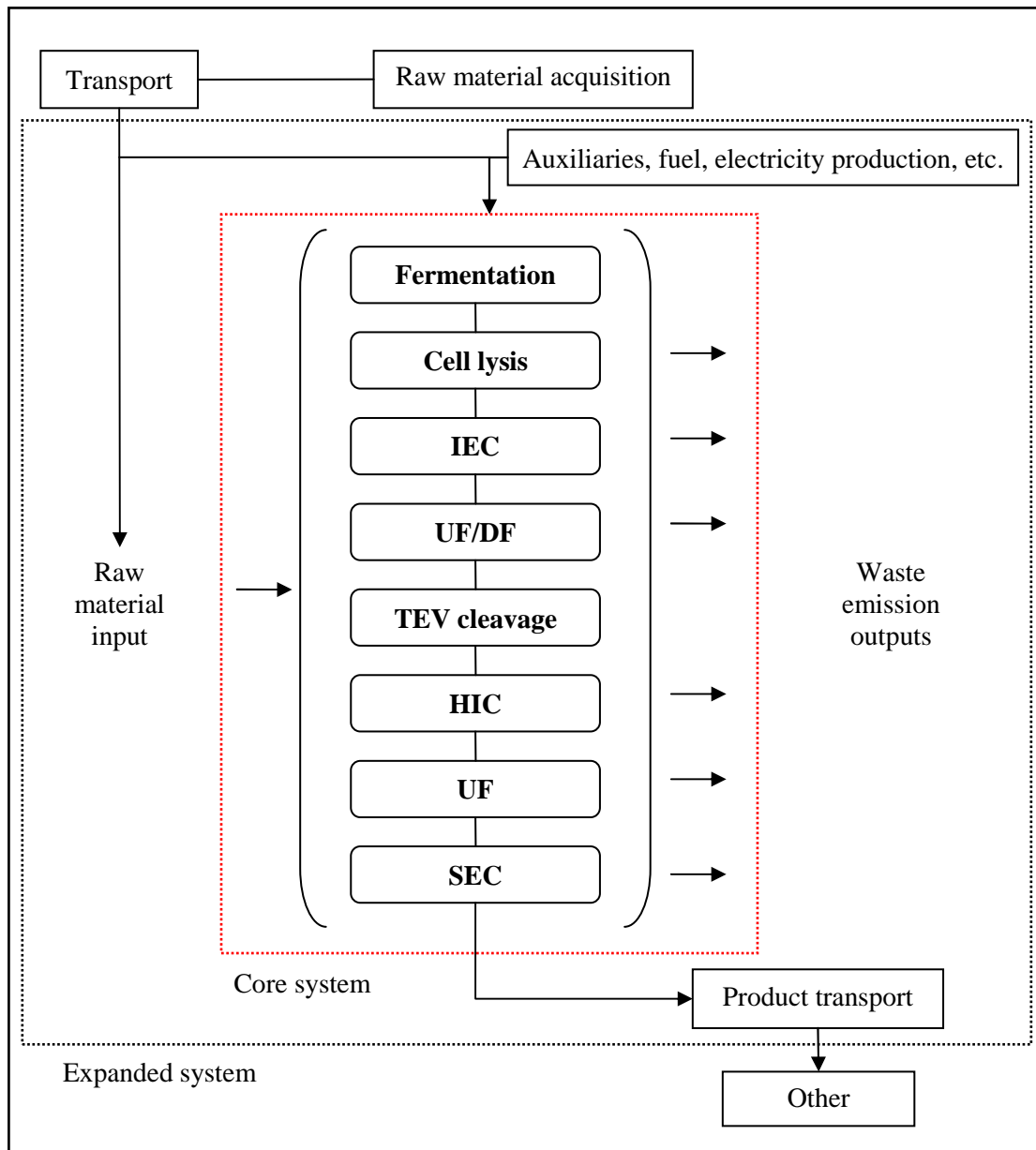


Figure 6.12 Schematic boundaries (gate-to-gate) for recombinant β -galactosidase production using the traditional multi-chromatographic purification scheme

Red dotted line (....): outlines core system; black dotted line (....): outlines expanded system; Fermentation: model E. coli strain employing unaltered TB media; IEC: ion exchange chromatography; UF/DF: ultrafiltration in combination with diafiltration; TEV cleavage: cleavage of His₆-tag from recombinant protein using TEV protease; HIC: hydrophobic interaction chromatography; UF: ultrafiltration; SEC: size exclusion chromatography

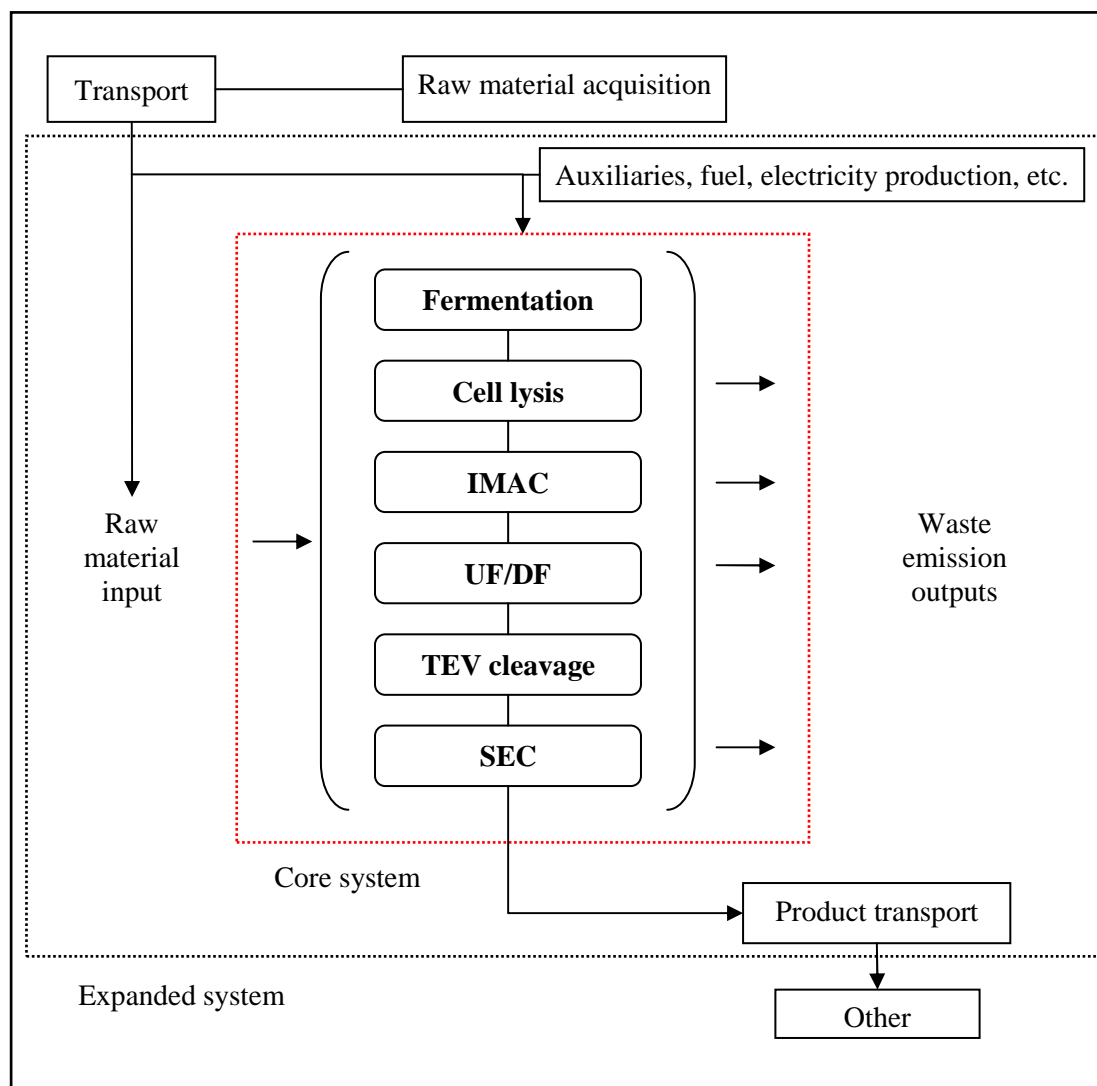


Figure 6.13 Schematic boundaries (gate-to-gate) for recombinant β -galactosidase production using the affinity-based purification strategy

Red dotted line (....): outlines core system; black dotted line (.....): outlines expanded system; Fermentation: model E. coli strain employing unaltered TB media; IMAC: immobilised metal affinity chromatography; UF/DF: ultrafiltration in combination with diafiltration; TEV cleavage: cleavage of His₆-tag from recombinant protein using TEV protease; SEC: size exclusion chromatography

6.3.1.2 Inventory analysis and impact assessment classification

Inventory data including TP raw material input and output waste emissions quantities from the recombinant protein production processes using the traditional multi-chromatographic and affinity-based purification schemes were determined from chapter three, section 3.4.2, chapter four, section 4.4.3 and chapter five, section 5.4.2. Solid and aqueous wastes as well as by-products generated during manufacture were identified by experimental methods as per chapters three, four and five unless otherwise stated.

The initial inventory and classification step of the EHS model during the application of the computer simulation for the production of the recombinant protein employing the traditional multi-chromatography purification scheme is illustrated in Figure 6.14. While Figure 6.15 shows the EHS simulation of the inventory and classification step for the recombinant β -galactosidase production using the affinity-based purification strategy. Total phosphorus (TP) was selected as the contaminant in the computer simulation as it is the pollutant investigated during this study. The computer simulation shows the individual core process steps together with their mass balance quantities (Q), which in this case are the total TP input and output at each stage entered manually. The TEV cleavage was not included as there is no waste stream produced that would leave the core system. Subsequently, the emission phases for each waste were established under the “Waste phase” heading and the phase effect factor (PE) was assigned to every waste. As previously outlined (section 5.4.1) the highest mass of TP emitted during both production processes was after the fermentation step, however, the outputs after the purification steps also included significant TP emissions mainly due to the use of sodium phosphate buffers. All aqueous waste streams were assigned a phase effect factor of 1 as these are considered to have a potential impact upon the environment. The solid waste generated during the cell lysis step were not further analysed in this research in an attempt to keep the study as accurate as possible since the focus was on liquid wastes.

Inventory: Determination of Quantity (Q)					
Select contaminant		Total Phosphorus (TP)			
Process Step	Process	Mass balance quantities		Waste phase	Phase effect factor (PE)
		Input (g)	Output (g) Q		
Process Step 1	Fermentation	0.274	0.213	Liquid	1
Process Step 2	Cell lysis	0.061	0.059	Solid	0
Process Step 3	Ion exchange chromatography (IEC)	0.002	0.002	Liquid	1
Process Step 4	Ultrafiltration/Diafiltration	0	0	Liquid	1
Process Step 5	Hydrophobic interaction chromatography (HIC)	0.182	0.153	Liquid	1
Process Step 6	Ultrafiltration	0.029	0.026	Liquid	1
Process Step 7	Size exclusion chromatography (SEC)	0.182	0.165	Liquid	1

Figure 6.14 Inventory and classification step for the recombinant β -galactosidase production process employing the traditional multi-chromatographic purification scheme analysed using EHS simulation model

Experimental determination of total P inputs and P waste emission outputs (Q) as well as assignment of phase effect factor (PE).

Inventory: Determination of Quantity (Q)					
Select contaminant	Total Phosphorus (TP)				
Mass balance quantities					
	Process	Input (g)	Output (g) Q	Waste phase	Phase effect factor (PE)
Process Step 1	Fermentation	0.274	0.213	Liquid	1
Process Step 2	Cell lysis	0.064	0.059	Solid	0
Process Step 3	Immobilised metal affinity chromatography (IMAC)	0.227	0.196	Liquid	1
Process Step 4	Ultrafiltration/Diafiltration	0.031	0.031	Liquid	1
Process Step 5	Size exclusion chromatography (SEC)	0.179	0.165	Liquid	1

Figure 6.15 Inventory and classification step for the recombinant β -galactosidase production process employing the affinity-based purification scheme analysed using EHS simulation model

Experimental determination of total P inputs and P waste emission outputs (Q) as well as assignment of phase effect factor (PE).

6.3.1.3 Characterisation

The next stage in the EHS model is the determination of the potential impact of the wastes emitted via the computation of the equivalency factor (EF) for the ecological impact categories as outlined in section 6.3.4. Total phosphorus impacts upon the environment in terms of one particular impact category: nutrient enrichment potential (NEP). As detailed in Figure 6.16, by employing the equation as shown in Table 6.1 the TP equivalency factor (EF) was calculated to be 1 g of phosphorus equivalent per gram of waste emitted. This study is specifically based on TP emissions and this pollutant does not impact significantly on the remaining environmental impact categories (AP, GWP, ODP, and POCP). Consequently, the equivalency factor for NEP is the only one computed during this case study.

For both recombinant protein production processes (using traditional multi-chromatographic and affinity-based purification) the specific pollutant selected is TP. Therefore, this characterisation step is the same for both determinations of the EHS and can be seen in Figure 6.16.

Characterisation: Determination of Equivalency Factors (EF)												
Environmental Impact Group 1	<u>Nutrient enrichment potential (NEP)</u> Calculation of NEP	<table border="1"> <tr> <td>Select contaminant formula</td> <td>$EF(P) = p * 30.97/MM$</td> <td rowspan="4">Equivalency factor (EF) g(P or N) g(substance)⁻¹</td> </tr> <tr> <td>Select N or P MM standard</td> <td>30.97</td> </tr> <tr> <td>Select p or v elements</td> <td>1</td> </tr> <tr> <td>Select MM of contaminant</td> <td>30.97</td> </tr> </table>	Select contaminant formula	$EF(P) = p * 30.97/MM$	Equivalency factor (EF) g(P or N) g(substance) ⁻¹	Select N or P MM standard	30.97	Select p or v elements	1	Select MM of contaminant	30.97	1.00
Select contaminant formula	$EF(P) = p * 30.97/MM$	Equivalency factor (EF) g(P or N) g(substance) ⁻¹										
Select N or P MM standard	30.97											
Select p or v elements	1											
Select MM of contaminant	30.97											
Environmental Impact Group 2	<u>Acidification potential (AP)</u> Calculation of AP											
Environmental Impact Group 3	<u>Global warming potential (GWP)</u> Calculation of GWP											
Environmental Impact Group 4	<u>Ozone depletion potential (ODP)</u> Calculation of ODP											
Environmental Impact Group 5	<u>Photochemical ozone creation potential (POCP)</u> Calculation of POCP											

Figure 6.16 Outline of characterisation step in the determination of EHS using simulation detailing specifically the computation of the equivalency factor (EF) for nutrient enrichment potential (NEP)

6.3.1.4 Determination of hazard score (HS)

The hazard score (HS) for each waste stream generated during the two production processes was determined as described in section 6.2.4. The computation results are detailed in Figure 6.17 for the process using the traditional multi-chromatographic purification and in Figure 6.18 employing the affinity-based purification scheme. Not surprisingly, waste streams with high TP quantities have high HS values while low TP amounts emitted in the waste streams results in a low HS value as the mass of TP in the waste emissions is directly involved in the calculation of the HS. In addition, if the phase effect factor for the pollutant in that particular emission phase was determined to be “0” than the HS will be 0, as seen in waste streams 2 (Figure 6.17 and 6.18). These are the solid wastes generated during cell lysis and as outlined in section 6.3.1.2 the PE is “0” for TP released in the solid emission phase.

Hazard Score Calculation: Determination of Hazard score (HS)			
$HS = \sum (EF * Q * PE)$			
Waste Stream 1	Q	0.21	Hazard Score 0.21
	PE	1.00	
	EF	1.00	
Waste Stream 2	Q	0.06	Hazard Score 0.00
	PE	0.00	
	EF	1.00	
Waste Stream 3	Q	0.00	Hazard Score 0.00
	PE	1.00	
	EF	1.00	
Waste Stream 4	Q	0.00	Hazard Score 0.00
	PE	1.00	
	EF	1.00	
Waste Stream 5	Q	0.15	Hazard Score 0.15
	PE	1.00	
	EF	1.00	
Waste Stream 6	Q	0.03	Hazard Score 0.03
	PE	1.00	
	EF	1.00	
Waste Stream 7	Q	0.17	Hazard Score 0.17
	PE	1.00	
	EF	1.00	

Figure 6.17 Outline showing hazard score (HS) calculation in the EHS simulation model for each of the waste streams generated during the recombinant protein production employing the traditional multi-chromatographic purification scheme

Hazard score calculations: Determination of Hazard score (HS)				
$HS = \sum (EF * Q * PE)$				
Waste Stream 1	Q	0.21		Hazard Score
	PE	1.00		0.21
	EF	1.00		
Waste Stream 2	Q	0.06		Hazard Score
	PE	0.00		0.00
	EF	1.00		
Waste Stream 3	Q	0.20		Hazard Score
	PE	1.00		0.20
	EF	1.00		
Waste Stream 4	Q	0.03		Hazard Score
	PE	1.00		0.03
	EF	1.00		
Waste Stream 5	Q	0.17		Hazard Score
	PE	1.00		0.17
	EF	1.00		

Figure 6.18 Outline showing hazard score (HS) calculation in the EHS simulation model for each of the waste streams generated during the recombinant protein production employing the affinity-based purification scheme

6.3.1.5 Valuation

The generation of the prioritisation factors for the weighting of the environmental impact groups is the next step of the EHS model and was carried out as described in section 6.2.5 (Figures 6.6-6.8). The results of the valuation stage for the process employing the traditional multi-chromatographic and the affinity-based purification scheme are shown in Figures 6.19 and 6.20, respectively. To maintain clarity, only the liquid waste streams generated during both production processes were further examined in the model. The production of recombinant β -galactosidase using the traditional multi-chromatographic purification scheme results in six aqueous waste streams (termed W_1 - W_6) as detailed in section 5.4.1.1, while employing the affinity-based purification strategy produces four liquid waste streams (entitled W_1 - W_4) as outlined in section 5.4.1.2.

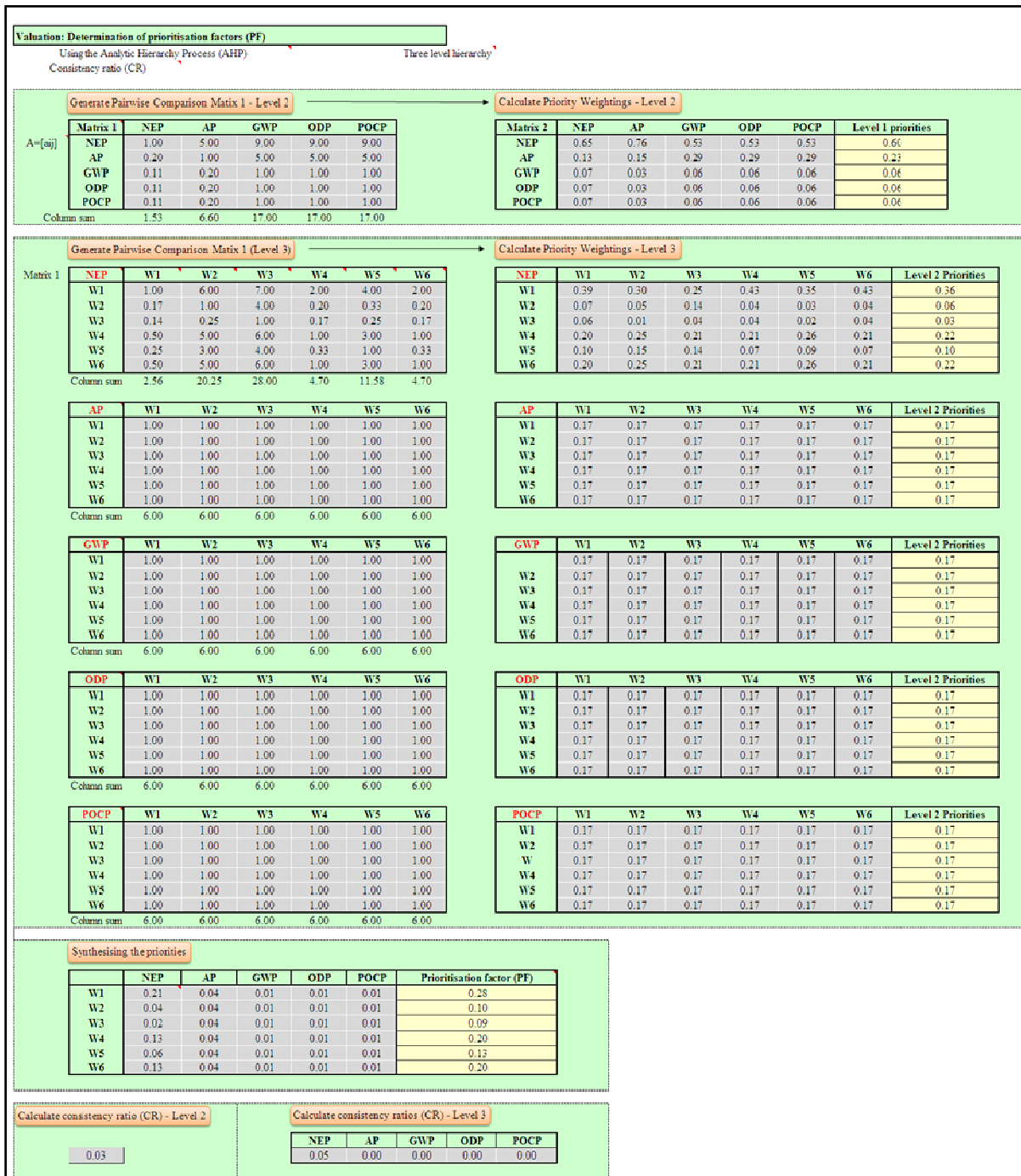


Figure 6.19 Overview of valuation step in the determination of the EHS for the recombinant protein production using the traditional multi-chromatographic purification scheme

Detailed determination of the prioritisation factor (PF) for each aqueous waste stream (W₁-W₆) generated during this production process. Consistency ratios for levels 1 and 2 were also calculated utilising the computer simulation.

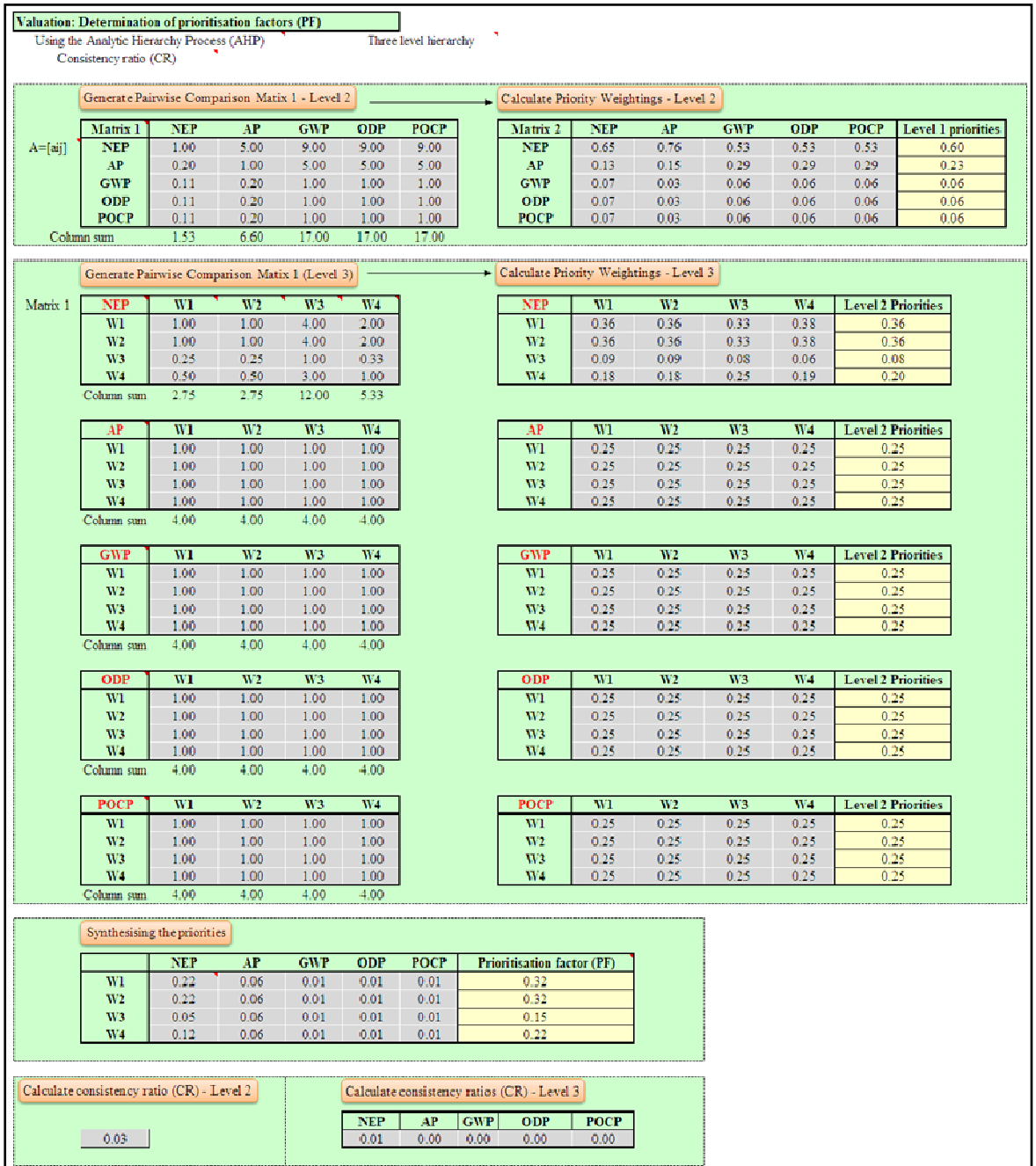


Figure 6.20 Overview of valuation step in the determination of the EHS for the recombinant protein production using the affinity-based purification scheme

Detailed determination of the prioritisation factor (PF) for each aqueous waste stream (W₁-W₄) generated during this production process. Consistency ratios for levels 1 and 2 were also calculated utilising the computer simulation.

The first phase in the valuation step requires a structural formulation of the problem into a hierarchy as described in section 6.2.5 and Figure 6.6 and in this case study, a three level hierarchy was designed. The goal of the AHP analysis (Level 1) was to evaluate the potential ecological impacts upon the environment from the total phosphorus emissions generated during the recombinant protein production process. The second level (criteria) consists of the five environmental impact categories necessary for ecological characterisation of the wastes and the last level (Level 3) makes up subcriteria alternatives to be examined which in this case study are the aqueous waste streams under investigation. Employing the traditional multi-chromatographic purification scheme during downstream processing results in six liquid wastes and consequently, there are six alternatives in Level 3 (see Figure 6.21). Since during the production process with the affinity-based purifications strategy only four liquid waste streams arise, only four alternatives were included in Level 3.

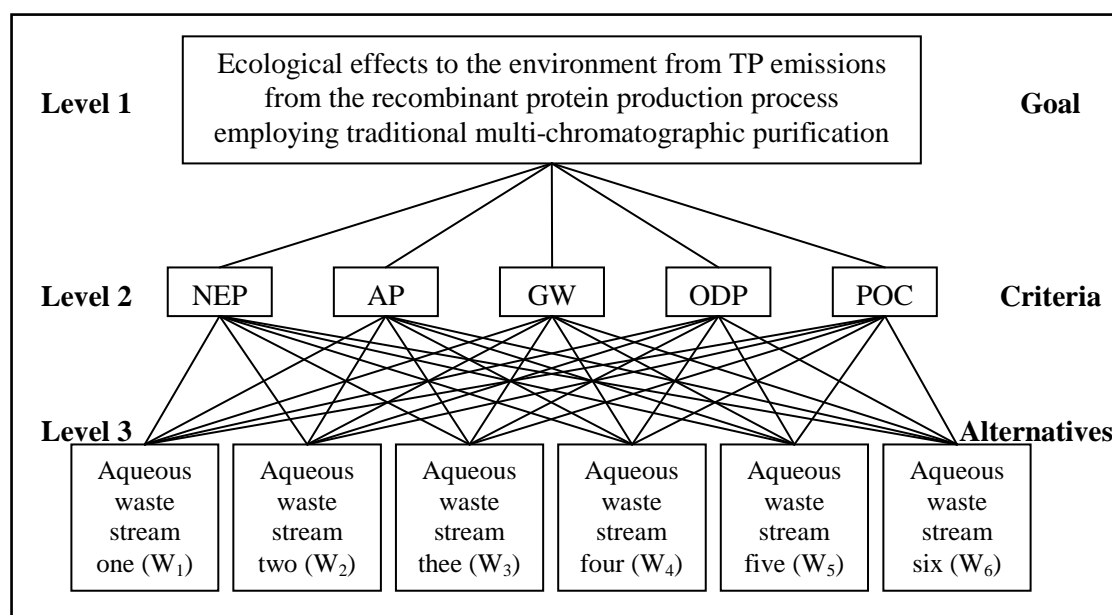


Figure 6.21 The hierarchy structure for the emissions from the recombinant β -galactosidase production process using the traditional multi-chromatographic purification scheme

The goal of the AHP analysis is shown in level 1 which is to determine the ecological effects to the environment from TP emissions from the recombinant protein production process employing the traditional multi-chromatographic purification scheme. Level 2 illustrates the five criteria necessary for ecological characterisation which are nutrient enrichment potential (NEP), acidification potential (AP), global warming potential (GWP), ozone depletion potential (ODP) and photochemical ozone creation potential (POCP). The alternatives to be evaluated in the hierarchy structure are presented in level 3 which are the aqueous waste streams generated during the production process (W_1 - W_6).

At this point, the relative importance of a set of criteria and alternatives were determined according to section 6.2.5 (phase 2). The pair-wise comparisons were evaluated by employing the standardised numerical comparison scale (Table 6.2) and the data was entered into the computer simulation which allowed subsequent calculation of the priority vectors. The matrices of the pair-wise comparisons between the criteria on Level 2 of the hierarchy with respect to the goal (Level 1) are shown in Table 6.3, as well as in Figure 6.19 and 6.20 for the two separate production processes. The rationale for the pair-wise comparisons were based on information obtained from the literature [Hauschild and Wenzel 1998; Wenzel *et al.* 2000].

Table 6.3 Pair-wise comparison matrix for ecological impact categories (Level 2) with respect to the goal (Level 1)

	NEP	AP	GWP	ODP	POCP	Priority vector
NEP	1.00	5.00	9.00	9.00	9.00	0.60
AP	0.20	1.00	5.00	5.00	5.00	0.23
GWP	0.11	0.20	1.00	1.00	1.00	0.06
ODP	0.11	0.20	1.00	1.00	1.00	0.06
POCP	0.11	0.20	1.00	1.00	1.00	0.06
Column sum	1.53	6.60	17.00	17.00	17.00	1.00

$CR = 0.03$, $RI = 1.12$;

NEP: nutrient enrichment potential, AP: acidification potential, GWP: global warming potential, ODP: ozone depletion potential, POCP: photochemical ozone creation potential

The priority vectors for the matrix were determined as described in section 6.2.5 and part (ii) in Figure 6.7. It involved dividing each pair-wise comparison entry by its column sum and averaging the resulting matrix over the rows. The priority vectors for Level 2 are 0.60, 0.23, 0.06, 0.06 and 0.06 as shown in Table 6.3 [Kablan 2004; Saaty and Vargas 2012]. The consistency ratio (CR) is calculated according to section 6.2.5 (and part (iii) in Figure 6.7) to ensure that the judgements are consistent. The first step in the computation of CR is the determination of λ_{\max} which is calculated by multiplying each pair-wise comparison value by its corresponding priority vector. In this case, the values of the first column in the matrix (Table 6.3) are multiplied by 0.600, the values of the second column of the matrix are multiplied

by 0.233 and so forth. Subsequently, the values of the resulting matrix were summed across and these values (3.277, 1.193, 0.281, 0.281 and 0.281) were divided by their corresponding priority vector and then averaged to give λ_{\max} . The steps for the λ_{\max} calculation are shown in Equations 6.1-6.3.

Determination of λ_{\max} :

$$\begin{bmatrix} 1.000 & 5.000 & 9.000 & 9.000 & 9.000 \\ 0.200 & 1.000 & 5.000 & 5.000 & 5.000 \\ 0.110 & 0.200 & 1.000 & 1.000 & 1.000 \\ 0.110 & 0.200 & 1.000 & 1.000 & 1.000 \\ 0.110 & 0.200 & 1.000 & 1.000 & 1.000 \end{bmatrix} \begin{bmatrix} 0.600 \\ 0.233 \\ 0.056 \\ 0.056 \\ 0.056 \end{bmatrix} \quad \text{(Equation 6.1)}$$

$$\begin{bmatrix} 0.600 & 1.165 & 0.504 & 0.504 & 0.504 \\ 0.120 & 0.233 & 0.280 & 0.280 & 0.280 \\ 0.066 & 0.047 & 0.056 & 0.056 & 0.056 \\ 0.066 & 0.047 & 0.056 & 0.056 & 0.056 \\ 0.066 & 0.047 & 0.056 & 0.056 & 0.056 \end{bmatrix} \begin{bmatrix} 3.277 \\ 1.193 \\ 0.281 \\ 0.281 \\ 0.281 \end{bmatrix} \quad \text{(Equation 6.2)}$$

Followed by:

$$\lambda_{\max} = \frac{\left(\frac{3.277}{0.600}\right) + \left(\frac{1.193}{0.233}\right) + \left(\frac{0.281}{0.056}\right) + \left(\frac{0.281}{0.056}\right) + \left(\frac{0.281}{0.056}\right)}{5} \quad \text{(Equation 6.3)}$$

$$\lambda_{\max} = 5.128$$

The consistency index (CI) can now be determined as outlined in section 6.2.5 and part (iii) in Figure 6.7.

$$CI = \frac{(\lambda_{\max} - n)}{(n - 1)} = \frac{(5.128 - 5)}{(5 - 1)} = 0.032 \quad \text{(Equation 6.4)}$$

Where n is the total number of objectives (size of matrix)

Finally, the consistency ratio (CR) was calculated as described in section 6.2.5 and part (iii) in Figure 6.7.

$$CR = \frac{CI}{RI} = \frac{0.032}{1.12} = \mathbf{0.029} \quad \text{(Equation 6.5)}$$

Where RI is the random consistency index (Table A, part (iii) in Figure 6.7), which is 1.12 for $n = 5$.

The consistency ratio for the pair-wise comparison of the ecological categories assessed (Table 6.3) was determined to be 0.03 as shown in Equation 6.5 and as this CR is below 0.1 the entries made in Table 6.3 are considered to be consistent [Saaty and Vargas 2012]. Reassessment of the judgements would have been necessary if the CR were ≥ 0.1 (section 6.2.5).

The next step involved the pair-wise comparison of the waste streams (alternatives, Level 3) with respect to the ecological impact categories (criteria, Level 2). These can be seen in Figures 6.19 and 6.20 which were obtained as part of the computer simulation of the EHS calculation for the recombinant protein production processes using the two different purification schemes. Again, only the aqueous waste streams were investigated further to maintain clarity. The liquid waste streams W_1 - W_6 (for the production process employing the traditional multi-chromatographic purification scheme) and W_1 - W_4 (for the manufacturing process using the affinity-based purification strategy) were compared and weighted in relation to their potential impact upon each ecological impact category investigated. The priority vectors, consistency and random consistency ratios for each of these matrices were determined as previously outlined and are also shown in Figures 6.19 and 6.20. As the consistency ratios for each of these pair-wise comparisons are well below 0.1, the judgements were considered to be consistent.

For the β -galactosidase production process with the traditional multi-chromatographic purification scheme the priority vectors with respect to the NEP criteria (see Figure 6.19) indicate that liquid waste stream one (W_1) has the highest probability to affect the nutrient enrichment as it has a priority vector value of 0.36 (36 %). This is a result of the large amount of TP remaining in the spent fermentation media (section 3.4.2) discharged in W_1 which can have a considerable impact on NEP. The aqueous waste streams four and six (W_4 and W_6) also have high importance values of 22 % for both due to the presence of phosphate salts in the buffers used during the protein purification steps. The remaining waste streams (W_2 ,

W_3 and W_5) have priority vector values of 0.06 (6 %), 0.03 (3 %) and 0.10 (10 %), respectively. These can be attributed to potential carryover of P (W_2 and W_3 , section 5.4.2.1) and low discharge quantities of P salts from the purification buffers (W_5 , section 5.4.2.1).

The priority vectors in relation to the NEP criteria for recombinant protein production process employing the affinity-based purification strategy show that aqueous waste streams one and two (W_1 and W_2) have an equal probability of 36 % of impacting upon the nutrient enrichment. This is due to the large amount of P containing spent media (W_1) and the phosphate salts employed in the IMAC purification buffer (W_2). Liquid waste stream four (W_4) arises after SEC and also contains high quantities of TP, again due to the P salts in the buffer, which results in an importance value of 20 % for W_4 . The lowest priority vector was observed for W_3 with a value of 0.08 (8 %) which can be attributed to carryover from the previous waste streams (section 5.4.2.2).

The pair-wise comparisons of the waste stream alternatives for the remaining ecological hazard groups (AP, GWP, ODP and POCP) are considered to have equal importance and were therefore assigned values of 1. Total phosphorus emissions are considered not to impact on these environmental groups [Hauschild and Wenzel 1998; Wenzel *et al.* 2000].

The priority vectors for each level of the hierarchy were determined to get an overall priority for each alternative which could then be used to calculate the prioritisation factor (section 6.2.5). The first step is the computation of the prioritisation factors of the alternatives (waste streams) at the lowest level of the hierarchy. This involves multiplication of the priority vectors of the waste stream alternatives with the priority vectors of the ecological impact categories (criteria). Taking the production process of recombinant β -galactosidase with the traditional multi-chromatographic purification scheme as an example it means multiplying the values of the first column by 0.60, multiplying the values of the second column by 0.23 and so forth (see Equation 6.6). During the final step the rows are summed across to give the prioritisation factor for TP emissions in each of the waste stream generated (Equation 6.7).

$$\begin{bmatrix} 0.356 & 0.167 & 0.167 & 0.167 & 0.167 \\ 0.062 & 0.167 & 0.167 & 0.167 & 0.167 \\ 0.033 & 0.167 & 0.167 & 0.167 & 0.167 \\ 0.224 & 0.167 & 0.167 & 0.167 & 0.167 \\ 0.103 & 0.167 & 0.167 & 0.167 & 0.167 \\ 0.224 & 0.167 & 0.167 & 0.167 & 0.167 \end{bmatrix} \begin{bmatrix} 0.600 \\ 0.230 \\ 0.056 \\ 0.056 \\ 0.056 \end{bmatrix} = \begin{bmatrix} 0.213 & 0.039 & 0.009 & 0.009 & 0.009 \\ 0.037 & 0.039 & 0.009 & 0.009 & 0.009 \\ 0.020 & 0.039 & 0.009 & 0.009 & 0.009 \\ 0.134 & 0.039 & 0.009 & 0.009 & 0.009 \\ 0.062 & 0.039 & 0.009 & 0.009 & 0.009 \\ 0.134 & 0.039 & 0.009 & 0.009 & 0.009 \end{bmatrix} \quad \text{(Equation 6.6)}$$

$$\begin{bmatrix} 0.213 & 0.039 & 0.009 & 0.009 & 0.009 \\ 0.037 & 0.039 & 0.009 & 0.009 & 0.009 \\ 0.020 & 0.039 & 0.009 & 0.009 & 0.009 \\ 0.134 & 0.039 & 0.009 & 0.009 & 0.009 \\ 0.062 & 0.039 & 0.009 & 0.009 & 0.009 \\ 0.134 & 0.039 & 0.009 & 0.009 & 0.009 \end{bmatrix} = \begin{bmatrix} 0.279 \\ 0.103 \\ 0.086 \\ 0.200 \\ 0.128 \\ 0.200 \end{bmatrix} \quad \text{(Equation 6.7)}$$

Figures 6.19 and 6.20 show overviews of these computations obtained via the computer simulation of the EHS model for both processes under investigation. The individual waste streams within each production process can now be compared evenly by employing the prioritisation factors. For the manufacturing of recombinant β -galactosidase using the traditional multi-chromatographic purification scheme during downstream processing, the prioritisation factors indicate that liquid waste stream one (W_1) has the greatest potential environmental impact (28 %) in terms of TP emissions. This is followed by aqueous waste streams four and six (W_4 and W_6) with prioritisation factors of 0.20 (20 %) and then W_5 (13 %), W_2 (10 %) and W_3 (9 %).

For the production process of the recombinant protein with the affinity-based purification strategy, the prioritisation factors for the liquid waste streams are as follows: 0.32 (32 %, W_1), 0.32 (32 %, W_2), 0.15 (15 %, W_3) and 0.22 (22 %, W_4). This indicates that aqueous waste streams one and two (W_1 and W_2) have both highest possibility to cause an environmental impact in terms of TP emissions, followed by W_4 and W_3 .

6.3.1.6 Calculation of ecological hazard score (EHS)

The final step in the EHS model computation procedure is the calculation of the ecological hazard score, which is achieved by multiplying the hazard score (HS)

from each waste stream with its relevant prioritisation factor as outlined in Figure 6.9, section 6.2.6. For the production process with the traditional multi-chromatographic purification scheme the HS values and prioritisation factors are shown in Figures 6.17 and 6.19, respectively, and the results for the EHS computation are illustrated in Figure 6.22.

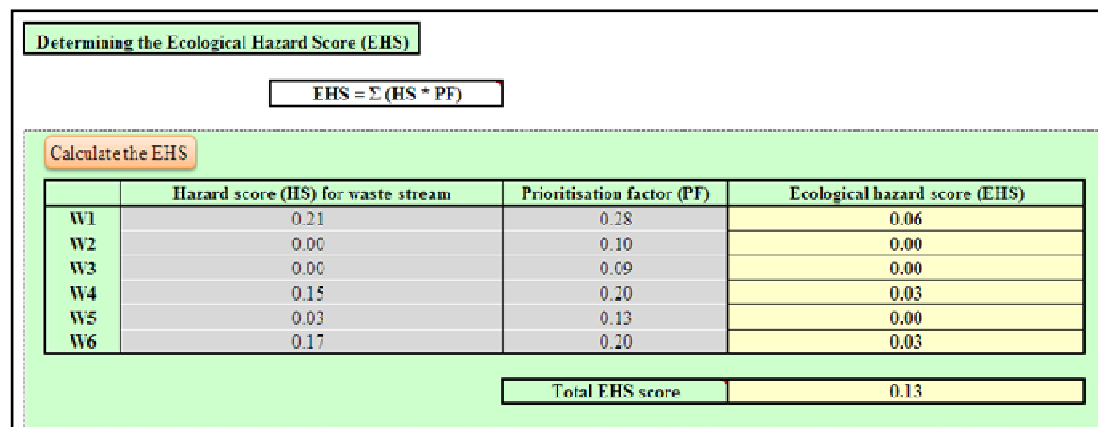


Figure 6.22 Overview of ecological hazard score (EHS) computation as determined for the recombinant protein production process employing the traditional multi-chromatographic purification scheme

The EHS for aqueous waste stream one (W_1) was 0.06 and for liquid waste streams four and six (W_4 and W_6) it was 0.03, while the remaining waste streams (W_2 , W_3 and W_5) had an EHS value of zero. This indicates that from the waste streams generated during the recombinant β -galactosidase production with the traditional multi-chromatographic purification scheme at laboratory-scale W_1 , W_4 and W_6 possibly impact upon the environment in terms of TP emissions. Due to the low (or zero) TP outputs in the remaining waste streams their corresponding ecological hazard scores were regarded to be insignificant. Decision makers can use these results to evaluate the individual unit operations for potential reductions to the environmental impact with respect to the particular pollutant investigated. For example, reductions of the P concentration in fermentation media may be considered to reduce the EHS of liquid waste stream one (W_1).

Figures 6.18 and 6.20 show the HS values for each waste stream and their relevant prioritisation factor, respectively, for the recombinant protein production process employing the affinity-based purification scheme. The data was then used to

calculate the ecological hazard scores according to Figure 6.9, section 6.2.6 and the results are presented in Figure 6.23. The findings show that the EHS values for aqueous waste streams one, two and four (W_1 , W_2 and W_4) are 0.07, 0.06 and 0.04, respectively, while W_3 has an EHS of zero. Therefore, the waste streams W_1 , W_2 and W_4 generated during the recombinant protein production with affinity-based purification at laboratory-scale can potentially impact upon the environment with respect to TP emissions.

Determining the Ecological Hazard Score (EHS)			
EHS = Σ (HS * PF)			
Calculate the EHS			
	Hazard score (HS) for waste stream	Prioritisation factor (PF)	Ecological hazard score (EHS)
W1	0.21	0.32	0.07
W2	0.20	0.32	0.06
W3	0.03	0.15	0.00
W4	0.17	0.22	0.04
Total EHS Score			0.17

Figure 6.23 Overview of ecological hazard score (EHS) computation as determined for the recombinant protein production process employing the affinity-based purification strategy

These results also demonstrate that the unit operations which generate these specific waste streams could require investigation for environmental impact reductions particularly with respect to phosphorus. For instance, reduction of the P concentration in fermentation media and replacing sodium phosphate salts with alternative buffering system may be considered to reduce the EHS values for liquid waste stream one, two and four (W_1 , W_2 and W_4).

The individual EHS values for the separate waste streams were summed to determine the total ecological hazard score (EHS) for each production process. The results, which can be seen in Figures 6.22 and 6.23, indicate that the production process for the recombinant model protein using the traditional multi-chromatographic purification scheme has a marginally lower overall potential impact on the environment with respect to TP emissions than the production process employing the affinity-based purification strategy. The difference between the two processes is the downstream processing set-up, i.e. the purification steps and therefore the volume and chemical composition of the waste streams emitted. While the results in chapter

four clearly highlighted a number of benefits when using the affinity-based purification scheme including reduced number of steps involved, increased product yield and reduced purification costs, the findings from the EHS computation evidently show that this production process has a marginally higher environmental impact in terms of TP emissions.

6.3.2 Case study two: Application of the EHS model to the production of recombinant protein production using affinity-based purification particular in terms of total phosphorus (TP) emissions

During this case study the EHS model was employed to evaluate the environmental impact of the waste streams emitted during the production of the recombinant β -galactosidase using the affinity-based purification process with the two fermentation media and buffering systems as described in chapter five of this thesis:

1. Fermentation of model *E. coli* with TB media and downstream processing employing the affinity-based purification strategy (as shown in Figure 5.2 in section 5.4.1.2)
2. Fermentation of model *E. coli* with 70 % P-reduced TB media and downstream processing employing the affinity-based purification strategy with the alternative buffering system (P-reduced production process; as described section 5.4.1.3).

Each of the production processes was assessed separately using the EHS model and the individual scores of each waste streams allowed an evaluation of the changes made (i.e. 70 % P-reduced TB media and alternative buffer) in terms of environmental performance.

The wastes arising were examined according to the same environmental impact groups (section 6.2.4): nutrient enrichment potential (NEP), acidification potential (AP), global warming potential (GWP), ozone depletion potential and photochemical ozone creation potential (POCP). The functional unit and system boundary were unchanged from the previous case study (section 6.3.1.1), i.e. functional unit was defined as grams of phosphorus equivalents per gram of waste ($\text{g(P-eq) g(waste)}^{-1}$) released to the environment in each of the waste streams for the two different processes. The same approach was used to determine the system boundaries and only

the core unit operations involved in the recombinant β -galactosidase production as illustrated in Figure 6.13 were investigated.

The steps in the model for EHS calculation are as outlined in case study one and the results for the recombinant production process employing the affinity-based purification scheme are detailed in section 6.3.1. Figures 6.24-6.27 illustrate the findings for the EHS computation, as obtained from the computer simulation, of the recombinant protein production using the affinity-based purification scheme with the 70 % P-reduced fermentation media and the alternative buffering system (P-reduced production process).

Inventory: Determination of Quantity (Q)					
Select contaminant	Total Phosphorus (TP)				
	Process	Mass balance quantities		Waste phase	Phase effect factor (PE)
		Input (g)	Output (g) Q		
Process Step 1	Fermentation	0.112	0.068	Liquid	1
Process Step 2	Cell lysis	0.024	0.023	Solid	0
Process Step 3	Immobilised metal affinity chromatography (IMAC)	0.001	0.001	Liquid	1
Process Step 4	Ultrafiltration/Diafiltration	0	0	Liquid	1
Process Step 5	Size exclusion chromatography (SEC)	0	0	Liquid	1

Figure 6.24 Inventory and classification step for the recombinant protein production process employing 70 % P-reduced TB media and the affinity-based purification scheme with the alternative buffering system analysed using EHS simulation model

Experimental determination of total P inputs and P waste emission outputs (Q) as well as assignment of phase effect factor (PE).

The TP waste outputs generated during this process are largely in the first aqueous waste stream (W_1) after fermentation due to the remaining nutrients in the spent media and the solid waste (SW) obtained after cell lysis due to the cellular debris. However, by reducing the P concentration in the fermentation media the TP outputs have significantly decreased as can be seen in Figure 6.24.

The characterisation step was carried out according to section 6.2.4 and is the same as in the previous case study resulting in an equivalency factor (EF) of 1 meaning 1 g of phosphorus equivalent per gram of waste emitted (see Figure 6.16).

The hazard scores (HS) for each of the waste streams generated for the P-reduced production process were calculated as described in section 6.2.4 and the results are

illustrated in Figure 6.25. The findings showed that aqueous waste stream one (W_1) had the highest HS value of 0.09 while the remaining waste streams had negligible hazard scores. Therefore, only W_1 may be a possible ecological hazard especially in terms of nutrient enrichment.

Hazard score calculations: Determination of Hazard score (HS)			
HS = Σ (EF * Q* PE)			
Waste Stream 1	Q	0.09	Hazard Score 0.09
	PE	1.00	
	EF	1.00	
Waste Stream 2	Q	0.02	Hazard Score 0.00
	PE	0.00	
	EF	1.00	
Waste Stream 3	Q	0.00	Hazard Score 0.00
	PE	1.00	
	EF	1.00	
Waste Stream 4	Q	0.00	Hazard Score 0.00
	PE	1.00	
	EF	1.00	
Waste Stream 5	Q	0.00	Hazard Score 0.00
	PE	1.00	
	EF	1.00	

Figure 6.25 Outline showing hazard score (HS) calculation is the EHS simulation model for each of the waste streams generated during the P-reduced recombinant protein production process

Valuation for the recombinant production process using affinity-based purification with 70 % P-reduced TB media and the alternative buffers was carried out as outlined in section 6.2.5.

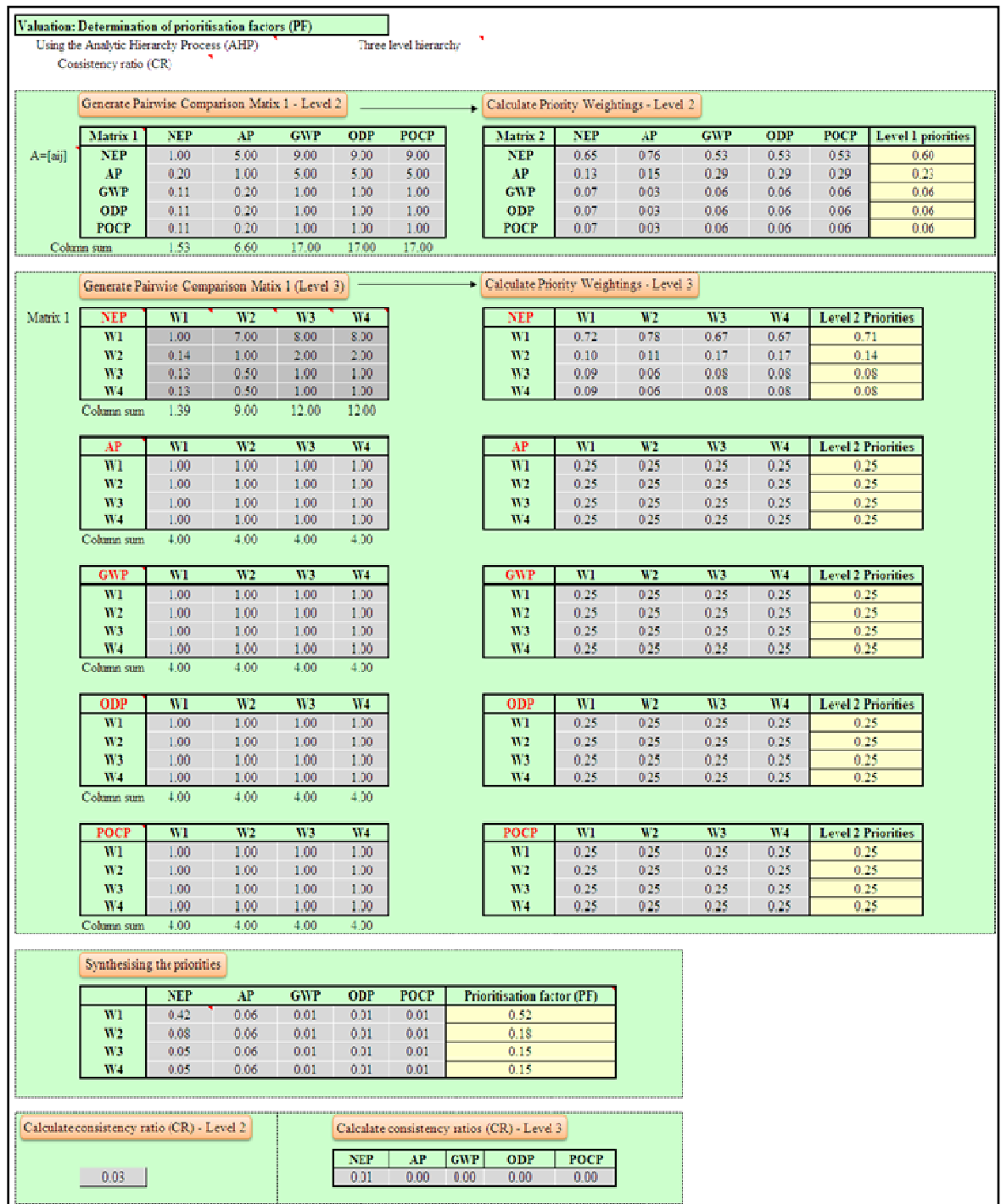


Figure 6.26 Overview of valuation step in the determination of the EHS for the recombinant protein production using affinity-based purification scheme with 70 % P-reduced TB media and the alternative buffers

Detailed determination of the prioritisation factor (PF) for each aqueous waste stream (W₁-W₄) generated during this production process. Consistency ratios for levels 1 and 2 were also calculated utilising the computer simulation.

The pair-wise comparison of the ecological hazard groups was unchanged and the comparisons for the liquid waste stream alternatives were re-evaluated with the new

TP output quantities. Figure 6.26 shows the results of the valuation step in the EHS computation for the P-reduced production process. The consistency ratios for all pair-wise comparison matrices are less than 0.1 indicating consistent judgements. The resulting prioritisation factors, as seen in Figure 6.26, indicate that waste stream one (W_1) has the highest potential to impact upon the environment in terms of TP emissions with a priority value of 0.52 (52 %). This is due to the residual P-containing compounds in the spent fermentation media. The remaining liquid waste streams W_2 , W_3 and W_4 have priority values of 0.18 (18 %), 0.15 (15 %) and 0.15 (15 %), respectively.

The EHS values for each waste stream were determined according to section 6.2.6 and the results for the P-reduced process are presented in Figure 6.27. Aqueous waste stream one (W_1) had an EHS value of 0.5 while the remaining waste streams reported a score of zero. This indicates that of all the waste streams arising from the production of the recombinant protein employing the P-reduced process at laboratory-scale only W_1 has a minimal potential to impact upon the environment in terms of total phosphorus emissions. The residual wastes have negligible or zero P outputs resulting in ecological hazard scores that can be deemed as insignificant. Consequently, these waste streams do not considerably impact upon the environment with respect to P emissions.

Determining the Ecological Hazard Score (EHS)			
$EHS = \sum (HS * PF)$			
Calculate the EHS			
	Hazard score (HS) for waste stream	Prioritisation factor (PF)	Ecological hazard score (EHS)
W1	0.09	0.52	0.05
W2	0.00	0.18	0.00
W3	0.00	0.15	0.00
W4	0.00	0.15	0.00
Total EHS Score			0.05

Figure 6.27 Overview of ecological hazard score (EHS) computation as determined for the recombinant protein production process employing the affinity-based purification strategy with 70 % P-reduced TB media and the alternative buffering systems

6.3.2.1 Comparison of P-reduced recombinant protein production process with standard process at laboratory-scale

Decision makers can evaluate changes made to the recombinant protein production process in terms of their environmental impacts with the help of the EHS model by comparing the waste streams generated. For instance, the waste stream and thereby the unit operation with the lower EHS value is regarded to have a less effect upon the ecological impact groups investigated in the model. This will allow design of a more environmentally friendly process with respect to a particular pollutant.

In this case study, the EHS values for two recombinant β -galactosidase production processes were determined and can now be compared to evaluate the environmental benefits from employing a 70 % P-reduced TB media during fermentation and substituting sodium phosphate salts with ACES during affinity-based purification. The results of the EHS value comparisons are shown in Table 6.4. The findings highlight that by reducing the P concentration in the fermentation media the environmental impact of the resulting waste stream (W_1) can be reduced.

Table 6.4 Ecological hazard score (EHS) for the four waste streams (W_1 - W_4) investigated for the standard and P-reduced recombinant protein production process with respect to total phosphorus emissions

	Standard process	P-reduced process
Aqueous waste stream one (W_1)	0.07	0.05
Aqueous waste stream two (W_2)	0.06	0
Aqueous waste stream three (W_3)	0	0
Aqueous waste stream four (W_4)	0.04	0
Total EHS score	0.17	0.05

Standard process: recombinant protein production using TB media and the affinity-based purification scheme

P-reduced process: recombinant protein production using 70 % P-reduced TB media and the affinity-based purification scheme with the alternative buffering system (ACES salts)

In addition, by replacing sodium phosphate salts with an alternative buffer not containing any P-compounds such as ACES the wastes generated during downstream processing exhibit a significantly reduced EHS. This in turn directly reflects their reduced potential to negatively impact upon the environment.

Overall, by employing the P-reduced production process over the standard process a reduction in the total ecological hazard score of 70 % could be achieved. This reduction is specifically in terms of nutrient enrichment potential (NEP), which can be defined as the impact on the ecosystem from substances containing phosphorus (or nitrogen) [Hauschild and Wenzel 1998; Wenzel *et al.* 2000].

6.4 Conclusion

This chapter described an analysis of the recombinant protein production processes, proposed throughout this study, in terms of their environmental impacts using the ecological hazard score (EHS) model. This allowed an overall environmental comparison of the processes but also highlighted potential opportunities for environmental improvements within each process.

Case study one (section 6.3.1), which was a comparison of the traditional multi-chromatographic and affinity-based purification systems, showed that while there is no major difference in environmental impact between the two processes, employing the traditional multi-chromatographic purification scheme results in a marginally “greener” process. This is due to the large amounts of TP discharged during the affinity-based purification of the recombinant protein recognising these waste streams as environmental danger areas. However, as previously shown the advantages of employing affinity-based chromatography are the reduced number of steps necessary to achieve purification and the increased product yield.

The analysis also identified that of all the individual unit operations in the production processes, the fermentation step generates a waste stream (W_1) with the greatest potential to negatively impact upon the environment in terms nutrient enrichment. In addition, it was observed that employing phosphate salt buffers during protein purification will also significantly add to the pollutant potential of the waste streams arising during recombinant protein production. Therefore, reductions of phosphorus in the fermentation media and reducing or replacing the P salts in the chromatography buffers were identified as environmental improvement opportunities for developing “greener” manufacturing processes.

These findings were further investigated in case study two (section 6.3.2) which evaluates the modifications to the recombinant protein production process using the affinity-based purification with respect to its environmental improvements. Significant environmental savings were achieved, i.e. a 70 % reduction in the total ecological hazard score, by employing the 70 % P-reduced TB media during fermentation and the alternative buffering system (ACES salts) during downstream processing.

Recognising these environmental “hot spots” during the early stages of the process development provides useful information relating to the decision for pilot plant- and industrial-scale. This in turn will allow the design of a production process with minimised environmental impacts.

Main findings and future work

Recombinant expression systems are regularly employed in the biopharmaceutical industry to generate therapeutic proteins and one of the most widely used expression hosts is *Escherichia coli*. The manufacturing of these products consists of the protein synthesis phase in a bioreactor/fermentation vessel (upstream processing) followed by purification and generation of the final product (downstream processing). This study investigated potential strategies to minimise the environmental burden associated with recombinant protein production in *E. coli*.

The initial part of this study employed standard molecular manipulation techniques to generate the *E. coli* strain MC1061 producing a recombinant protein with an N-terminal His₆-tag. The protein selected is *E. coli* β -galactosidase, an enzyme that has been studied extensively and reviewed in detail. Simple standard activity assays employing chromogenic, fluorogenic or luminogenic substrates for the enzyme are well established. Recombinant protein expression was optimised to ensure high concentration of soluble, functional β -galactosidase. The engineered model *E. coli* strain was subsequently used throughout this study to investigate the potential to reduce phosphorus employed in recombinant protein production.

Stepwise minimisation of the phosphorus content in complex (Terrific broth) and semi-defined media (minimal media 9 supplemented with yeast extract) was carried out during fermentation of the model *E. coli* strain and the effects on biomass yield and the quantity and quality of the recombinant β -galactosidase evaluated. The findings indicate that reductions of up to 70 % of exogenous phosphate salts added to both media were possible without adversely impacting on the cellular mass obtained or the recombinant protein produced. High concentrations of P employed during fermentation operations have the potential for significant environmental burdens such as resource depletion, added pressure on the waste water treatment facilities and the increased risk of eutrophication. The proposed reductions in P inputs during large-scale recombinant protein production can result in considerable phosphorus savings and an overall cleaner manufacturing process.

The next stage of this research entailed the development of two protein isolation strategies for the recombinant β -galactosidase enzyme. One purification scheme was based on a traditional multi-chromatographic approach and the other employing

affinity-based chromatography using the polyhistidine protein tag. While both schemes employed accomplished similar final purification of the protein, utilising affinity chromatography resulted in a reduced number of steps necessary to achieve product homogeneity. This in turn led to an increase in yield as well as a 60 % decrease in downstream processing costs. The reduced buffer consumption for this purification scheme decreases the potential environmental impacts generated during downstream processing by reducing the total P, ammonia and chemical oxygen demand inputs. In addition, analyses to reduce the P wastes associated with protein purification were carried out by minimising or replacing the phosphate salts in the chromatographic solvents with alternative buffering compounds such as ACES salts. The findings in this study highlighted great potential to further reduce or even eliminate the P outputs in the waste streams that arise during protein purification without negative product consequences. These combined environmental and economic benefits showed real industrial potential for affinity-based purification with P-optimised buffering systems.

The mass balance simulations of the recombinant β -galactosidase production process identified environmental “hot spots” which may require monitoring and treatment. Of particular concern is the spent fermentation media due to its high chemical oxygen demand and total P concentrations which can be attributed to the residual unused nutrients and biological components present. The pollution load of the remaining waste streams that arise during the manufacturing process depend on the chemical composition of the buffers employed during protein purification. This study showed that by reducing/eliminating the P inputs during fermentation and downstream processing the P outputs in the generated waste streams can be reduced significantly. This in turn can reduce production as well as waste treatment costs and contribute to designing an environmentally friendlier manufacturing process.

Environmental modelling was used throughout this study to illustrate the environmental impacts associated with recombinant protein manufacturing. The specific model employed is the ecological hazard score (EHS) model which allowed the evaluation of the whole β -galactosidase production process as well as establish specific waste streams or unit operations for environmental improvement opportunities. The analysis showed that the waste stream generated during

fermentation, i.e. the spent media, has the highest potential to negatively impact upon the environment in terms of nutrient enrichment. In addition, buffers with phosphate salts used during downstream processing can also add to the pollution potential of the recombinant protein production process. However, the findings also showed that by reducing the P concentration in the fermentation media and replacing P compounds in the purification buffers with alternative salts can lead to considerable environmental savings resulting in a “greener” β -galactosidase manufacturing process.

In order to establish the likely effect of P minimisation during large-scale production as well as allow control of media pH during product formation fermentation trials may need to be carried out in bioreactors. These microbial cell fermentors include ports to monitor pH, temperature and often specific nutrient concentrations while also facilitating heat and oxygen transfer as well as the addition of acid/base for pH adjustments. During the scale-up experimentations with P-reduced fermentation media other process parameters such as exact nutrient composition in the defined media, dissolved oxygen demand and growth and expression temperatures may also be varied in bioreactors to ensure high biomass and hence product yield.

Additional future work utilising fermentation vessels may also include P minimisation for all three types of fermentation processes: batch, fed-batch and continuous. This research focused on P reductions in media employed for shake-flask experiments which previous studies have shown can be related to batch fermentation. However, to ensure wide range large-scale applications of P-reduced fermentation media it may be necessary to show that there is potential to minimise P inputs in fed-batch or continuous fermentation processes.

This study has provided concepts to reduce the potential P outputs associated with biopharmaceutical manufacturing using *E. coli* to produce a recombinant protein. It is likely that these approaches can also be applied to recombinant protein production systems utilising other expression systems. For example, many therapeutics are being generated with the use of animal cell cultures and there is definite scope to evaluate the possibility to reduce P during upstream and downstream processing.

Bibliography

- Aguilar-Lasserre, A. A., Bautista Bautista, M. A., Ponsich, A. and González Huerta, M. A. (2009) 'An AHP-based decision-making tool for the solution of multiproduct batch plant design problem under imprecise demand', *Computers and Operations Research*, 36(3), 711-736.
- Al-Dasoqi, N., Mason, A., Alkhaddar, R. and Al-Shamma'a, A. (2011) *Use of Sensors in Wastewater Quality Monitoring—A Review of Available Technologies*, Conference, 22-26 May 2011, Location: American Society of Civil Engineers, 3379-3388.
- Altschul, S. F., Madden, T. L., Schaeffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997) 'Gapped BLAST and PSI-BLAST: a new generation of protein database search programs', *Nucleic Acids Research*, 25, 3389-3402.
- Aristidou, A. A., San, K.-Y. and Bennett, G. N. (1999) 'Improvement of biomass yield and recombinant gene expression in *Escherichia coli* by using fructose as the primary carbon source', *Biotechnology Progress*, 15(1), 140-145.
- Arnau, J., Lauritzen, C., Petersen, G. E. and Pedersen, J. (2006) 'Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins', *Protein Expression and Purification*, 48, 1-13.
- Atlas, R. M. (2004) *Handbook of Microbiological Media*, 3rd ed., Boca Raton: CRC Press.
- Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G. and Struhl, K., eds. (1988) *Current Protocols in Molecular Biology*, New York: John Wiley & Sons.
- Azapagic, A. (1999) 'Life cycle assessment and its application to process selection, design and optimisation', *Chemical Engineering Journal*, 73(1), 1-21.
- Azapagic, A. and Clift, R. (1999a) 'The application of life cycle assessment to process optimisation', *Computers and Chemical Engineering*, 23(10), 1509-1526.
- Azapagic, A. and Clift, R. (1999b) 'Life cycle assessment and multiobjective optimisation', *Journal of Cleaner Production*, 7(2), 135-143.
- Bachmann, B. J. (1972) 'Pedigrees of some mutant strains of *Escherichia coli* K-12', *Bacteriological Reviews*, 36(4), 525-557.
- Baneyx, F. (1999) 'Recombinant protein expression in *Escherichia coli*', *Current Opinion in Biotechnology*, 10(5), 411-421.
- Baneyx, F. (2004) *Protein Expression Technologies: Current Status and Future Trends*, Wymondham: Horizon Bioscience.
- Behme, S. (2009) *Manufacturing of pharmaceutical protein: from technology to economy*, Weinheim: Wiley-Blackwell.
- Berg, U., Donnert, D., Weidler, P. G., Kaschka, E., Knoll, G. and Nüesch, R. (2006) 'Phosphorus removal and recovery from wastewater by tobermorite-seeded crystallisation of calcium phosphate', *Water Science and Technology*, 53(3), 131-138.

- Bhander, G. S., Hauschild, M. and McAloone, T. (2003) 'Implementing life cycle assessment in product development', *Environmental Progress*, 22(4), 255-267.
- Biwer, A. P., Griffith, S. and Cooney, C. (2005a) 'Uncertainty analysis of penicillin V production using Monte Carlo simulation', *Biotechnology and Bioengineering*, 90(2), 167-179.
- Biwer, A. P., Zuber, P. T., Zelic, B., Gerharz, T., Bellmann, K. J. and Heinzle, E. (2005b) 'Modeling and Analysis of a New Process for Pyruvate Production', *Industrial and Engineering Chemistry Research*, 44(9), 3124-3133.
- Blattner, F. R., Plunkett, G., III, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B. and Shao, Y. (1997) 'The Complete Genome Sequence of *Escherichia coli* K-12', *Science*, 277(5331), 1453-1462.
- Bonetta, L. (2006) 'Protein purification: Fast forward', *Nature*, 439(7079), 1017-1021.
- Bradford, M. M. (1976) 'A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding', *Analytical Biochemistry*, 72(1-2), 248-254.
- Brase, C. H. and Brase, C. P. (2009) *Understandable Statistics: Concepts and Methods*, 9th ed., Boston: Brooks/Cole Publishing Co.
- Brickman, E., Silhavy, T. J., Bassford, P. J., Jr, Shuman, H. A. and Beckwith, J. R. (1979) 'Sites within gene *lacZ* of *Escherichia coli* for formation of active hybrid β -galactosidase molecules', *Journal of Bacteriology*, 139(1), 13-18.
- Brunet, R., Guillén-Gosálbez, G., Pérez-Correa, J. R., Caballero, J. A. and Jiménez, L. (2012) 'Hybrid simulation-optimization based approach for the optimal design of single-product biotechnological processes', *Computers and Chemical Engineering*, 37(0), 125-135.
- Cappucino, J. G. and Sherman, N. (2005) *Microbiology. A Laboratory Manual*, 7th ed., San Francisco: Benjamin Cummings.
- Carta, G. and Jungbauer, A. (2010) 'Downstream Processing of Biotechnology Products' in *Protein Chromatography: Process Development and Scale-Up*, Weinheim: Wiley-VCH, 1-55.
- Carty, G., O'Leary, G. and Meaney, B. (1997) *Wastewater Treatment Manuals. Primary, Secondary and Tertiary Treatment*, 1/97/400, Ardacavan, Wexford.: Environmental Protection Agency, Ireland.
- Chakraborti, S., Sani, R. K., Banerjee, U. C. and Sobti, R. C. (2000) 'Purification and characterization of a novel β -galactosidase from *Bacillus* sp MTCC 3088', *Journal of Industrial Microbiology and Biotechnology*, 24(1), 58-63.
- Chen, R. (2012) 'Bacterial expression systems for recombinant protein production: *E. coli* and beyond', *Biotechnology Advances*, 30(5), 1102-1107.

- Choi, A. C. K., Kaebernick, H. and Lai, W. H. (1997) 'Manufacturing processes modelling for environmental impact assessment', *Journal of Materials Processing Technology*, 70(1-3), 231-238.
- Choi, J. H., Keum, K. C. and Lee, S. Y. (2006) 'Production of recombinant proteins by high cell density culture of *Escherichia coli*', *Chemical Engineering Science*, 61(3), 876-885.
- Chorus, I. and Bartam, J., eds. (1999) *Toxic cyanobacteria in water: A guide to their public health consequences, monitoring and management*, 1st ed., London: E & FN Spon on behalf of World Health Organisation.
- Ciambrone, D. F. (1997) *Environmental Life Cycle Analysis*, Boca Raton: CRC Lewis Publishers.
- Cikankowitz, A. and Laforest, V. (2013) 'Using BAT performance as an evaluation method of techniques', *Journal of Cleaner Production*, 42(0), 141-158.
- Cinar, A. (2003) *Batch Fermentation: Modeling Monitoring and Control*, Basel: Marcel Dekker Incorporated.
- Cliffe, F. E. (2010) *A study of selected environmental issues related to the production of plasmid DNA for application in gene therapy*, unpublished thesis (Ph.D. Thesis), University of Limerick.
- Cliffe, F. E., Walsh, G. and O'Dwyer, T. F. (2010) 'Utilisation of phosphorus nutrient content in industrial scale plasmid DNA production: a waste minimisation study', *Journal of Cleaner Production*, 18(10-11), 1066-1072.
- Cohen, S. N., Chang, A. C. Y. and Hsu, L. (1972) 'Nonchromosomal Antibiotic Resistance in Bacteria: Genetic Transformation of *Escherichia coli* by R-Factor DNA', *Proceedings of the National Academy of Sciences of the United States of America*, 69(8), 2110-2114.
- Corchero, J. L., Viaplana, E., Benito, A. and Villaverde, A. (1996) 'The position of the heterologous domain can influence the solubility and proteolysis of β -galactosidase fusion proteins in *E. coli*', *Journal of Biotechnology*, 48(3), 191-200.
- Corchero, J. L. and Villaverde, A. (1999) 'Tolerance of *Escherichia coli* β -galactosidase C-terminus to different-sized fusions', *Biotechnology and Bioengineering*, 64(6), 644-649.
- Cordell, D., Drangert, J.-O. and White, S. (2009) 'The story of phosphorus: Global food security and food for thought', *Global Environmental Change*, 19(2), 292-305.
- Cordell, D., Rosemarin, A., Schröder, J. J. and Smit, A. L. (2011) 'Towards global phosphorus security: A systems framework for phosphorus recovery and reuse options', *Chemosphere*, 84(6), 747-758.
- Cramer, S. M. and Holstein, M. A. (2011) 'Downstream bioprocessing: recent advances and future promise', *Current Opinion in Chemical Engineering*, 1(1), 27-37.

- Craven, G. R., Steers, E. and Anfinsen, C. B. (1965) 'Purification, composition and molecular weight of β -galactosidase of *Escherichia coli* K12 ', *Journal of Biological Chemistry*, 240(6), 2468-2477.
- Crespo, A. L., Spencer, T. A., Nekl, E., Pusztai-Carey, M., Moar, W. J. and Sigfried, B. D. (2008) 'Comparison and Validation of methods to quantify Cry1Ab toxin from *Bacillus thuringiensis* for standardization of insect bioassays', *Applied and Environmental Microbiology*, 74(1), 130-135.
- Crouzet, P., Leonard, J., Nixon, S., Rees, Y., Parr, W., Laffon, L., Bogestrand, J., Kristensen, P., Lallana, C., Izzo, G., Bokn, T., Bak, J., Lack, T. and Thyssen, N. (1999) *Nutrients in European ecosystems*, no 4, Copenhagen: European Environment Agency.
- Csathó, P., Sisák, I., Radimsky, L., Lushaj, S., Spiegel, H., Nikolova, M. T., Nikolov, N., Čermák, P., Klir, J., Astover, A., Karklins, A., Lazauskas, S., Kopyński, J., Hera, C., Dumitru, E., Manojlovic, M., Bogdanović, D., Torma, S., Leskošek, M. and Khristenko, A. (2007) 'Agriculture as a source of phosphorus causing eutrophication in Central and Eastern Europe', *Soil Use and Management*, 23, 36-56.
- Cupples, C., Miller, J. and Huber, R. (1990) 'Determination of the roles of Glu-461 in β -galactosidase (*Escherichia coli*) using site-specific mutagenesis', *Journal of Biological Chemistry*, 265(10), 5512-5518.
- Cupples, C. G. and Miller, J. H. (1988) 'Effects of Amino Acid Substitutions at the Active Site in *Escherichia coli* β -Galactosidase', *Genetics*, 120(3), 637-644.
- D'Souza, J. I. and Killedar, S. G. (2008) *Biotechnology And Fermentation Process*, Pune: Nirali Prakashan.
- Dandres, T., Gaudreault, C., Tirado-Seco, P. and Samson, R. (2012) 'Macroanalysis of the economic and environmental impacts of a 2005–2025 European Union bioenergy policy using the GTAP model and life cycle assessment', *Renewable and Sustainable Energy Reviews*, 16(2), 1180-1192.
- Das, D., Allen, T. M. and Suresh, M. R. (2005) 'Comparative evaluation of two purification methods of anti-CD19-c-myc-His6-Cys scFv', *Protein Expression and Purification*, 39(2), 199-208.
- Davies, J. E. and Demain, A. L., eds. (1999) *Manual of industrial microbiology and biotechnology*, 2nd ed., Washington D.C.: ASM Press.
- Dawson, R. M., Elliott, D. C., Elliott, W. H. and Jones, K. M. (1986) 'pH, buffers, and physiological media' in *Data for Biochemical Research*, 3rd ed., Oxford: Oxford University Press, 417-448.
- Demain, A. L. (2000) 'Microbial biotechnology', *Trends in Biotechnology*, 18(1), 26-31.
- Demain, A. L. (2007) 'The business of biotechnology', *Industrial Biotechnology*, 3(3), 269-283.
- Demain, A. L. and Vaishnav, P. (2009) 'Production of recombinant proteins by microbes and higher organisms', *Biotechnology Advances*, 27(3), 297-306.

- Deutscher, M. P., ed. (1990) *Guide to Protein Purification*, 1st ed., London: Academic Press Inc.
- Dowd, J., Cosentino, G. and Chao, S.-B. (2007) 'New Technologies in Biopharmaceutical Upstream Processing' in Dutton, R. L. and Scharer, J. M., eds., *Advanced Technologies in Biopharmaceutical Processing*, 1st ed., Oxford: Blackwell Publishing, 97-129.
- Dugdale, M. L., Vance, M. L., Wheatley, R. W., Driedger, M. R., Nibber, A., Tran, A. and Huber, R. E. (2010) 'Importance of Arg-599 of β -galactosidase (*Escherichia coli*) as an anchor for the open conformations of Phe-601 and the active-site loop', *Biochemistry and Cell Biology*, 88(6), 969-979.
- Dunn, P. J., Wells, A. S. and Williams, M. T., eds. (2010) *Green Chemistry in the Pharmaceutical Industry*, Weinheim: Wiley-VCH.
- Dupont, R. R., Theodore, L. and Ganesan, K., eds. (2000) *Pollution Prevention: The Waste Management Approach for the 21st Century*, Boca Raton: Lewis Publishers.
- Eaton, A. D. and Franson, M. A. H., eds. (1995) *Standard Methods for the examination of water and wastewater*, 19th ed., Washington, DC: American Health Association; American Water Works Association; Water Environment Federation.
- EC (2012) 'The IPPC Directive', [online], available: <http://ec.europa.eu/environment/air/pollutants/stationary/ippc/summary.htm> [accessed 30/11/2012].
- Edwards, J. (1997) 'Large-scale column chromatography - a GMP manufacturing perspective' in Goldberg, E., ed. *Handbook of downstream processing*, London: Blackie Academic & Professional, 167-184.
- EEA (2013) 'The European Pollutant Release and Transfer Register (E-PRTR)', [online], available: <http://prtr.ec.europa.eu/Home.aspx> [accessed 09Jul13].
- Eiteman, M. A. and Altman, E. (2006) 'Overcoming acetate in *Escherichia coli* recombinant protein fermentations', *Trends in Biotechnology*, 24(11), 530-536.
- El-Mansi, M., Bryce, C. F. A., Demain, A. L. and Allman, A. R. (2007) *Fermentation Microbiology And Biotechnology*, 2nd ed., Boca Raton: Taylor & Francis Group.
- EMA (1999) *Specifications: Test Procedures and Acceptance Criteria for Biotechnological/Biological Products*, CPMP/ICH/365/96, London: EMA.
- EMA (2013) 'European Medicines Agency', [online], available: http://www.ema.europa.eu/ema/index.jsp?curl=/pages/home/Home_Page.jsp&jsenab ed=true [accessed Jun2013].
- EMBL (2012) 'Bacterial Expression Vectors', [online], available: http://www.embl.de/pepcore/pepcore_services/strains_vectors/vectors/bacterial_expression_vectors/popup_bacterial_expression_vectors/ [accessed 22Aug2013].

- EPA (2008) 'BAT Guidance Note on Best Available Techniques for Pharmaceutical and Other Speciality Organic Chemicals', [online], available: <https://www.epa.ie/pubs/advice/bat/BAT%20Guidance%20Note%20Pesticides%20Pharmaceuticals%20&%20Speciality%20Organic%20Chemicals.pdf> [accessed].
- EPA (2010) 'Phosphorus Regulations', [online], available: <http://www.epa.ie/whatwedo/enforce/pa/phosphorus/> [accessed 15Apr10].
- EPA (2013) 'Integrated Pollution Prevention Control (IPPC) Licensing search', [online], available: <http://www.epa.ie/terminalfour/ippc/index.jsp?disclaimer=yes&Submit=Continue> [accessed 19Sep13].
- EUROPA (2011) 'Integration pollution prevention and control: IPPC Directive', [online], available: http://europa.eu/legislation_summaries/environment/waste_management/128045_en.htm [accessed 10Jul13].
- Farid, S. S. (2007) 'Process economics of industrial monoclonal antibody manufacture', *Journal of Chromatography B*, 848(1), 8-18.
- FDA (2013a) *Guidance for Industry. Immunogenicity assessment for therapeutic protein products*, UCM338856, Rockville: FDA.
- FDA (2013b) 'U.S. Food and Drug Administration', [online], available: <http://www.fda.gov/> [accessed Jun2013].
- Flickinger, M. C. and Drew, S. W., eds. (1999) *Encyclopedia of Bioprocess Technology*, New York: Wiley.
- Fong, B. A., Wu, W.-Y. and Wood, D. W. (2010) 'The potential role of self-cleaving purification tags in commercial-scale processes', *Trends in Biotechnology*, 28(5), 272-279.
- Forciniti, D. (2008) *Industrial Bioseparations: Principles and Practice*, Oxford: Blackwell Publishing Ltd.
- Forster, C. F. (2003) 'Phosphorus Removal' in *Wastewater Treatment and Technology*, Reston: Thomas Telford, 168-178.
- Fowler, A. V. and Zabin, I. (1977) 'The amino acid sequence of β -galactosidase of *Escherichia coli*', *Proceedings of the National Academy of Sciences of the United States of America*, 74(4), 1507-1510.
- Fowler, A. V. and Zabin, I. (1978) 'Amino acid sequence of β -galactosidase. XI. Peptide ordering procedures and the complete sequence', *Journal of Biological Chemistry*, 253(15), 5521-5525.
- Fowler, A. V. and Zabin, I. (1983) 'Purification, structure, and properties of hybrid β -galactosidase proteins', *Journal of Biological Chemistry*, 258(23), 14354-14358.

- Francoeur, S. N., Biggs, B. J. F., Smith, R. A. and Lowe, R. L. (1999) 'Nutrient Limitation of Algal Biomass Accrual in Streams: Seasonal Patterns and a Comparison of Methods', *Journal of the North American Benthological Society*, 18(2), 242-260.
- Fu, X. Y., Wei, D. Z. and Tong, W. Y. (2006) 'Effect of yeast extract on the expression of thioredoxin-human parathyroid hormone from recombinant *Escherichia coli*', *Journal of Chemical Technology and Biotechnology*, 81(12), 1866-1871.
- Gallagher, S. R. and Desjardins, P. R. (2008) 'Quantitation of DNA and RNA with Absorption and Fluorescence Spectroscopy', *Current Protocols in Protein Science*, 52, A.4K.1-A.4K.21.
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. and Bairoch, A. (2005) 'Protein Identification and Analysis Tools on the ExPASy server' in Walker, J. M., ed. *The Proteomics Protocol Handbook*, New York: Humana Press, 572-607.
- GE Healthcare (2007) *Recombinant Protein Purification Handbook*, Uppsala: GE Healthcare Bio-Sciences AB.
- GE Healthcare (2010a) *Gel Filtration. Principles and Methods*, Uppsala, Sweden: GE Healthcare Bio-Sciences AB.
- GE Healthcare (2010b) *Strategies for Protein Purification*, Uppsala, Sweden: GE Healthcare Bio-Sciences AB.
- GE Healthcare (2013) *Ion Exchange Chromatography and Chromatofocusing. Principles and Methods*, Uppsala: GE Healthcare Bio-Sciences AB.
- Gebler, J., Aebersold, R. and Withers, S. (1992) 'Glu-537, not Glu-461, is the nucleophile in the active site of (*lac Z*) β -galactosidase from *Escherichia coli*', *Journal of Biological Chemistry*, 267(16), 11126-11130.
- Gill, D. (2009) 'Protein Pharmaceuticals: Discovery and Preclinical Development' in Guzmán, C. and Feuerstein, G., eds., *Pharmaceutical Biotechnology. Advances in experimental Medicine and Biology*, New York: Springer Science and Business Media LLC., 28-36.
- Glick, B. R., Pasternak, J. J. and Patten, C. L. (2010) 'Large-Scale Production of Proteins from Recombinant Microorganisms' in *Molecular Biotechnology: Principles and Applications of Recombinant DNA*, 4th ed., Washington, D.C.: ASM Press, 685-724.
- Grant, S. G., Jessee, J., Bloom, F. R. and Hanahan, D. (1990) 'Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants', *Proceedings of the National Academy of Sciences*, 87(12), 4645-4649.
- Gräslund, S., Nordlund, P., Weigelt, J., Bray, J., Gileadi, O. and *et al.* (2008) 'Protein Production and Purification', *Nature Methods*, 5(2), 135-146.
- Gräslund, T., Lundin, G., Uhlén, M., Nygren, P.-Å. and Hober, S. (2000) 'Charge engineering of a protein domain to allow efficient ion-exchange recovery', *Protein Engineering*, 13(10), 703-709.

- Haes, H. U., Jolliet, O., Finnveden, G., Hauschild, M., Krewitt, W. and Müller-Wenk, R. (1999) 'Best available practice regarding impact categories and category indicators in life cycle impact assessment', *The International Journal of Life Cycle Assessment*, 4(2), 66-74.
- Ham, T. S., Lee, S. K., Keasling, J. D. and Arkin, A. P. (2006) 'A tightly regulated inducible expression system utilizing the fim inversion recombination switch', *Biotechnology and Bioengineering*, 94(1), 1-4.
- Hammer, M. J. and Hammer, M. J. J. (2004) *Water and Wastewater Technology*, 5th ed., New Jersey: Pearson Prentice Hall.
- Han, M.-J., Lee, S. Y., Koh, S.-T., Noh, S.-G. and Han, W. H. (2010) 'Biotechnological applications of microbial proteomes', *Journal of Biotechnology*, 145(4), 341-349.
- Hanrahan, G., Salmassi, T. M., Khachikian, C. S. and Foster, K. L. (2005) 'Reduced inorganic phosphorus in the natural environment: significance, speciation and determination', *Talanta*, 66(2), 435-444.
- Hardy, K. (1986) *Bacterial Plasmids*, 2nd ed., Birkshire: Springer.
- Hauschild, M. and Wenzel, H. (1998) *Environmental Assessment of Products: Scientific background*, London: Chapman & Hall.
- Heinzle, E., Biber, A. P. and Cooney, C. L. (2007) *Development of Sustainable Bioprocesses: Modeling and Assessment*, Chichester: Wiley.
- Henderson, P. J. F. (1993) 'Statistical analysis of enzyme kinetic data' in Eienthal, R. and Danson, M. J., eds., *Enzyme Assays - A Practical Approach*, Oxford: Oxford University Press, 277-316.
- Hertwich, E. G., Pease, W. S. and Koshland, C. P. (1997) 'Evaluating the environmental impact of products and production processes: a comparison of six methods', *Science of The Total Environment*, 196(1), 13-29.
- Hitzfeld, B. C., Höger, S. J. and Dietrich, D. R. (2000) 'Cyanobacterial toxins: removal during drinking water treatment, and human risk assessment', *Environmental Health Perspectives*, 108 (suppl 1), 113-122.
- Ho, R. J. Y. and Gibaldi, M. (2004) *Biotechnology and Biopharmaceuticals: Transforming Proteins and Genes into Drugs*, Wiley.
- Ho, S. V. (2010) 'Environmental Considerations in Biologics Manufacturing' in Dunn, P. J., Wells, A. S. and Williams, M. T., eds., *Green Chemistry in the Pharmaceutical Industry*, Weinheim: Wiley-VCH, 311-331.
- Holba, M., Skorvan, O., Marsalkova, E. and Marsalek, B. (2012) *Phosphorus removal from wastewater via environmentally friendly technologies*, Conference, 23rd-25th Oct 2012, Location.
- Horne, R. E., Grant, T. and Verghese, K. L. (2009) *Life Cycle Assessment: Principles, Practice, and Prospects*, Collingwood: CSIRO Publishing.

- Hortsch, R. and Weuster-Botz, D. (2011) 'Growth and recombinant protein expression with *Escherichia coli* in different batch cultivation media', *Applied Microbiology and Biotechnology*, 90(1), 69-76.
- Hu, A. S. L., Wolfe, R. G. and Reithel, F. J. (1959) 'The preparation and purification of β -galactosidase from *Escherichia coli*, ML 308', *Archives of Biochemistry and Biophysics*, 81(2), 500-507.
- Hubbuck, J. and Kula, M.-R. (2007) 'Isolation and Purification of Biotechnological Products', *Journal of Non-Equilibrium Thermodynamics*, 32(2), 99-127.
- Hui, I. K., Li, C. P. and Lau, H. C. W. (2003) 'Hierarchical environmental impact evaluation of a process in printed circuit board manufacturing', *International Journal of Production Research*, 41(6), 1149-1165.
- Hunt, I. (2005) 'From gene to protein: a review of new and enabling technologies for multi-parallel protein expression', *Protein Expression and Purification*, 40(1), 1-22.
- IMARC (2012) *Global Biopharmaceutical Market Report and Forecast: (2012-2017)*, International Market Analysis Research & Consulting (IMARC), 2243296.
- Jacob, F. and Monod, J. (1961) 'Genetic regulatory mechanisms in the synthesis of proteins', *Journal of Molecular Biology*, 3(3), 318-356.
- Jacob, L. R. (1999) 'Hydrophobic Interaction Chromatography' in Kastner, M., ed. *Protein Liquid Chromatography*, Amsterdam: Elsevier, 235-269.
- Jacobson, R. H. and Matthews, B. W. (1992) 'Crystallization of β -galactosidase from *Escherichia coli*', *Journal of Molecular Biology*, 223(4), 1177-1182.
- Jacobson, R. H., Zhang, X. J., DuBose, R. F. and Matthews, B. W. (1994) 'Three-dimensional structure of β -galactosidase from *E. coli*', *Nature*, 369(6483), 761-766.
- Jana, S. and Deb, J. K. (2005) 'Strategies for efficient production of heterologous proteins in *Escherichia coli*', *Applied Microbiology and Biotechnology*, 67(3), 289-298.
- Jancewicz, L. J., Wheatley, R. W., Sutendra, G., Lee, M., Fraser, M. E. and Huber, R. E. (2012) 'Ser-796 of β -galactosidase (*Escherichia coli*) plays a key role in maintaining a balance between the opened and closed conformations of the catalytically important active site loop', *Archives of Biochemistry and Biophysics*, 517(2), 111-122.
- Janson, J. C. (2012) *Protein Purification: Principles, High Resolution Methods, and Applications*, New Jersey: John Wiley & Sons Inc.
- Jensen, K. F. (1993) 'The *Escherichia coli* K-12 "wild types" W3110 and MG1655 have an rph frameshift mutation that leads to pyrimidine starvation due to low pyrE expression levels', *Journal of Bacteriology*, 175(11), 3401-3407.
- Josic, D., Kova, S. and ccaron (2008) 'Application of proteomics in biotechnology - Microbial proteomics', *Biotechnology Journal*, 3(4), 496-509.

- Juers, D. H., Hakda, S., Matthews, B. W. and Huber, R. E. (2003) 'Structural Basis for the Altered Activity of Gly794 Variants of *Escherichia coli* β -Galactosidase', *Biochemistry*, 42(46), 13505-13511.
- Juers, D. H., Heightman, T. D., Vasella, A., McCarter, J. D., Mackenzie, L., Withers, S. G. and Matthews, B. W. (2001) 'A structural view of the action of *Escherichia coli* (*lacZ*) β -galactosidase', *Biochemistry*, 40(49), 14781-14794.
- Juers, D. H., Jacobson, R. H., Wigley, D., Zhang, X. J., Huber, R. E., Tronrud, D. E. and Matthews, B. W. (2000) 'High resolution refinement of β -galactosidase in a new crystal form reveals multiple metal-binding sites and provides a structural basis for alpha-complementation', *Protein Science*, 9(9), 1685-1699.
- Junker, B. H. (2004) 'Scale-up methodologies for *Escherichia coli* and yeast fermentation processes', *Journal of Bioscience and Bioengineering*, 97(6), 347-364.
- Junker, B. H. (2010) 'Minimizing the Environmental Footprint of Bioprocesses. Part 2: Evaluation of Wastewater, Electricity, and Air Emissions', *BioProcess International*, 8(9), 36-46.
- Jurado, E., Camacho, F., Luzon, G. and Vicaria, J. M. (2002) 'A new kinetic model proposed for enzymatic hydrolysis of lactose by a β -galactosidase from *Kluyveromyces fragilis*', *Enzyme and Microbial Technology*, 31(3), 300-309.
- Kablan, M. M. (2004) 'Decision support for energy conservation promotion:: an analytic hierarchy process approach', *Energy Policy*, 32(10), 1151-1158.
- Kalnins, A., Otto, K., Ruether, U. and Mueller-Hill, B. (1983) 'Sequence of the *lacZ* gene of *Escherichia coli*', *EMBO Journal*, 2(4), 593-597.
- Kamionka, M. (2011) 'Engineering of Therapeutic Proteins Production in *Escherichia coli*', *Current Pharmaceutical Biotechnology*, 12(2), 268-274.
- Kastner, M. (1999) 'Immobilized Metal Ion Affinity Chromatography' in Kastner, M., ed. *Protein liquid chromatography*, Amsterdam: Elsevier, 301-383.
- Kensy, F., Zang, E., Faulhammer, C., Tan, R.-K. and Buchs, J. (2009) 'Validation of a high-throughput fermentation system based on online monitoring of biomass and fluorescence in continuously shaken microtiter plates', *Microbial Cell Factories*, 8(1), 31-48.
- Khan, F. A. and Ansari, A. A. (2005) 'Eutrophication: An Ecological Vision', *The Botanical Review*, 71(4), 449-82.
- Khan, F. I., Natrajan, B. R. and Revathi, P. (2001) 'GreenPro: a new methodology for cleaner and greener process design', *Journal of Loss Prevention in the Process Industries*, 14(4), 307-328.
- Knochel, J. P. (2006) 'Phosphorus' in Shils, M. E., Shike, M., Ross, A. C., Caballero, B. and Cousins, R. J., eds., *Modern Nutrition in Health and Diseases*, 10th ed., Baltimore: Lippincott Williams & Wilkins, 211-222.

- Knorre, W. A., Deckwer, W. D., Korz, D., Pohl, H. D., Riesenberg, D., Ross, A., Sanders, E. and Schulz, V. (1991) 'High-cell-density fermentation of recombinant *Escherichia coli* with computer-controlled optimal-growth rate', *Annals of the New York Academy of Sciences*, 646, 300-306.
- Kolaj, O., Spada, S., Robin, S. and Wall, J. G. (2009) 'Use of folding modulators to improve heterologous protein production in *Escherichia coli*', *Microbial Cell Factories*, 8(1), 9-26.
- Kopetzki, E., Lehnert, K. and Buckel, P. (1994) 'Enzymes in diagnostics: achievements and possibilities of recombinant DNA technology', *Clinical Chemistry*, 40(5), 688-704.
- Korz, D. J., Rinas, U., Hellmuth, K., Sanders, E. A. and Deckwer, W. D. (1995) 'Simple fed-batch technique for high cell density cultivation of *Escherichia coli*', *Journal of Biotechnology*, 39(1), 59-65.
- Kosinski, M. J. and Bailey, J. E. (1991) 'Temperature and induction effects on the degradation rate of an abnormal β -galactosidase in *Escherichia coli*', *Journal of Biotechnology*, 18(1-2), 55-68.
- Krause, M., Ukkonen, K., Haataja, T., Ruottinen, M., Glumoff, T., Neubauer, A., Neubauer, P. and Vasala, A. (2010) 'A novel fed-batch based cultivation method provides high cell-density and improves yield of soluble recombinant proteins in shaken cultures', *Microbial Cell Factories*, 9(1), 11-22.
- Kuby, S. A. and Lardy, H. A. (1953) 'Purification and kinetics of β -galactosidase from *Escherichia coli*, Strain K-12', *Journal of the American Chemical Society*, 75(4), 890-896.
- Kuhn, D., Kholiq, M. A., Heinzle, E., Buhler, B. and Schmid, A. (2010) 'Intensification and economic and ecological assessment of a biocatalytic oxyfunctionalization process', *Green Chemistry*, 12(5), 815-827.
- Kulikova, T., Akhtar, R., Aldebert, P., Althorpe, N., Andersson, M., Baldwin, A., Bates, K., Bhattacharyya, S., Bower, L., Browne, P., Castro, M., Cochrane, G., Duggan, K., Eberhardt, R., Faruque, N., Hoad, G., Kanz, C., Lee, C., Leinonen, R., Lin, Q., Lombard, V., Lopez, R., Lorenc, D., McWilliam, H., Mukherjee, G., Nardone, F., Pastor, M. P. G., Plaister, S., Sobhany, S., Stoehr, P., Vaughan, R., Wu, D., Zhu, W. and Apweiler, R. (2007) 'EMBL Nucleotide Sequence Database in 2006', *Nucleic Acids Research*, 35(suppl_1), D16-20.
- Kunitz, M. (1945) 'Crystallization of a trypsin inhibitor from soybean', *Science*, 101(2635), 668-669.
- Kuo, W.-H. and Chase, H. (2011) 'Exploiting the interactions between poly-histidine fusion tags and immobilized metal ions', *Biotechnology Letters*, 33(6), 1075-1084.
- Laemmli, U. K. (1970) 'Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4', *Nature*, 227(5259), 680-685.
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J.

- and Higgins, D. G. (2007) 'Clustal W and Clustal X version 2.0', *Bioinformatics*, 23(21), 2947-2948.
- Lauritzen, C., Nygaard, G. E., Pedersen, J., Arnau, J., Römer, U. and Schäfer, F. (2013) 'Reliable, Large-Scale Cleavage of Tags from Affinity-Purified Biopharmaceuticals' in *Modern Biopharmaceuticals*, Weinheim: Wiley-VCH Verlag GmbH & Co. KGaA, 539-557.
- Lawrence, J. V. and Maier, S. (1977) 'Correction for the inherent error in optical density readings', *Applied and Environmental Microbiology*, 33(2), 482-484.
- Lederberg, J. (1950) 'The β -D-galactosidase of *Escherichia coli*, strain K-12 ', *Journal of Bacteriology*, 60(4), 381-392.
- Lee, J. J., O'Callaghan, P. and Allen, D. (1995) 'Critical review of life cycle analysis and assessment techniques and their application to commercial activities', *Resources, Conservation and Recycling*, 13(1), 37-56.
- Lee, S. Y. (1996) 'High cell-density culture of *Escherichia coli*', *Trends in Biotechnology*, 14(3), 98-105.
- Li, C., Zhang, X., Zhang, S. and Suzuki, K. (2009) 'Environmentally conscious design of chemical processes and products: Multi-optimization method', *Chemical Engineering Research and Design*, 87(2), 233-243.
- Li, C. P. and Hui, I. K. (2001) 'Environmental Impact Evaluation Model for Industrial Processes', *Environmental Management*, 27(5), 729-737.
- Li, X., Robbins, J. W., Jr. and Taylor, K. B. (1990) 'The production of recombinant beta-galactosidase in *Escherichia coli* in yeast extract enriched medium', *Journal of Industrial Microbiology*, 5(2-3), 85-93.
- Li, Y. (2011) 'Self-cleaving fusion tags for recombinant protein production', *Biotechnology Letters*, 33(5), 869-881.
- Lilie, H., Schwarz, E. and Rudolph, R. (1998) 'Advances in refolding of proteins produced in *E. coli*', *Current Opinion in Biotechnology*, 9(5), 497-501.
- Link, A. J., Robison, K. and Church, G. M. (1997) 'Comparing the predicted and observed properties of proteins encoded in the genome of *Escherichia coli* K-12', *ELECTROPHORESIS*, 18(8), 1259-1313.
- Linninger, A. A., Stephanopoulos, E., Ali, S. A., Han, C. and Stephanopoulos, G. (1995) 'Generation and assessment of batch processes with ecological considerations', *Computers and Chemical Engineering*, 19, Supplement 1(0), 7-13.
- Lischke, H. H., Brandes, L., Wu, X. and Schügerl, K. (1993) 'Influence of acetate on the growth of recombinant *Escherichia coli* JM103 and product formation', *Bioprocess and Biosystems Engineering*, 9(4), 155-157.

- Losen, M., Frölich, B., Pohl, M. and Büchs, J. (2004) 'Effect of oxygen limitation and medium composition on *Escherichia coli* fermentation in shake-flask cultures', *Biotechnology Progress*, 20(4), 1062-1068.
- Lowe, C. R., Lowe, A. R. and Gupta, G. (2001) 'New developments in affinity chromatography with potential application in the production of biopharmaceuticals', *Journal of Biochemical and Biophysical Methods*, 49(1-3), 561-574.
- Lozano, S., Iribarren, D., Moreira, M. T. and Feijoo, G. (2010) 'Environmental impact efficiency in mussel cultivation', *Resources, Conservation and Recycling*, 54(12), 1269-1277.
- Luli, G. W. and Strohl, W. R. (1990) 'Comparison of growth, acetate production, and acetate inhibition of *Escherichia coli* strains in batch and fed-batch fermentations', *Applied and Environmental Microbiology*, 56(4), 1004-1011.
- Macauley-Patrick, S. and Finn, B. (2008) 'Modes of Fermenter Operation' in McNeil, B. and Harvey, L., eds., *Practical Fermentation Technology*, Chichester: Wiley, 69-96.
- Mackey, K. R. M. and Paytan, A. (2009) 'Phosphorus Cycle' in Schaechter, M., ed. *Encyclopedia of Microbiology*, Oxford: Elsevier Science, 322-334.
- Madigan, M. T., Martinko, J. M. and Parker, J. (2009) *Brock Biology of Microorganisms*, 12th ed., San Francisco: Pearson Education Inc.
- Maestro, B., Velasco, I., Castillejo, I., Arévalo-Rodríguez, M., Cebolla, Á. and Sanz, J. M. (2008) 'Affinity partitioning of proteins tagged with choline-binding modules in aqueous two-phase systems', *Journal of Chromatography A*, 1208(1-2), 189-196.
- Maher, W. and Woo, L. (1998) 'Procedures for the storage and digestion of natural waters for the determination of filterable reactive phosphorus, total filterable phosphorus and total phosphorus', *Analytica Chimica Acta*, 375(1-2), 5-47.
- Manahan, S. E. (2005) *Environmental Chemistry*, 8th ed., Boca Raton: CRC Press LLC.
- Marchetti, J. M. (2011) 'The effect of economic variables over a biodiesel production plant', *Energy Conversion and Management*, 52(10), 3227-3233.
- MarketLine (2012) *Biotechnology: Global Industry Guide*, Dublin: Research and Markets, 41522.
- Marvuglia, A., Benetto, E., Rege, S. and Jury, C. (2013) 'Modelling approaches for consequential life-cycle assessment (C-LCA) of bioenergy: Critical review and proposed framework for biogas production', *Renewable and Sustainable Energy Reviews*, 25(0), 768-781.
- Marx, C. K., Hertel, T. C. and Pietzsch, M. (2008) 'Purification and activation of a recombinant histidine-tagged pro-transglutaminase after soluble expression in *Escherichia coli* and partial characterization of the active enzyme', *Enzyme and Microbial Technology*, 42(7), 568-575.

- Massey, M. S., Ippolito, J. A., Davis, J. G. and Sheffield, R. E. (2010) 'Macroscopic and microscopic variation in recovered magnesium phosphate materials: Implications for phosphorus removal processes and product re-use', *Bioresource Technology*, 101(3), 877-885.
- Matthews, B. W. (2005) 'The structure of *E. coli* β -galactosidase', *Comptes Rendus Biologies*, 328(6), 549-556.
- Mayyas, A. T., Qattawi, A., Mayyas, A. R. and Omar, M. A. (2012) 'Life cycle assessment-based selection for a sustainable lightweight body-in-white design', *Energy*, 39(1), 412-425.
- McKinney, J., Knappskog, P. M., Pereira, J., Ekern, T., Toska, K., Kuitert, B. B., Levine, D., Gronenborn, A. M., Martinez, A. and Haavik, J. (2004) 'Expression and purification of human tryptophan hydroxylase from *Escherichia coli* and *Pichia pastoris*', *Protein Expression and Purification*, 33(2), 185-194.
- McKnight, J. L. and Fried, V. A. (1981) 'Limited proteolysis. Early steps in the processing of large premature termination fragments of β -galactosidase in *Escherichia coli*', *Journal of Biological Chemistry*, 256(18), 9652-9661.
- Mendes, M. R., Aramaki, T. and Hanaki, K. (2003) 'Assessment of the environmental impact of management measures for the biodegradable fraction of municipal solid waste in São Paulo City', *Waste Management*, 23(5), 403-409.
- Mitraki, A. and King, J. (1989) 'Protein Folding Intermediates and Inclusion Body Formation', *Nature Biotechnology*, 7(7), 690-697.
- Morton, S. C. and Edwards, M. (2005) 'Reduced Phosphorus Compounds in the Environment', *Critical Reviews in Environmental Science and Technology*, 35(4), 333-364.
- Müller, K. M., Gempeler, M. R., Scheiwe, M.-W. and Zeugin, B. T. (1996) 'Quality assurance for biopharmaceuticals: An overview of regulations, methods and problems', *Pharmaceutica Acta Helvetiae*, 71(6), 421-438.
- Musa, N. S. and Ahmad, W. A. (2010) 'Chemical Oxygen Demand reduction in industrial wastewater using locally isolated bacteria', *Malaysian Journal of Fundamental and Applied Sciences*, 6(2), 88-92.
- Nagle, P. C., Nicita, C. A., Gerdes, L. A. and Schmeichel, C. J. (2008) 'Characteristics of and Trends in the Late-stage Biopharmaceutical Pipeline', *The American Journal of Managed Care*, 14(4), 226-229.
- Nakano, S., Fujimoto, M., Hara, H. and Sugimoto, N. (1999) 'Nucleic acid duplex stability: influence of base composition on cation effects', *Nucleic Acids Research*, 27(14), 2957-2965.
- Nancib, N., Branlant, C. and Boudrant, J. (1991) 'Metabolic roles of peptone and yeast extract for the culture of a recombinant strain of *Escherichia coli*', *Journal of Industrial Microbiology and Biotechnology*, 8(3), 165-169.

- Nelson, D. L. and Cox, M. M. (2005) *Lehninger Principles of Biochemistry*, 4th ed., New York: W. H. Freeman and Company.
- New, A. P., Freitas dos Santos, L. M., Lo Biundo, G. and Spicq, A. (2000) 'Analytical techniques used for monitoring the biodegradation of fluorinated compounds in waste streams from pharmaceutical production', *Journal of Chromatography A*, 889(1-2), 177-184.
- Nichols, E., Gavina, J., McLeod, R. and Craig, D. (2007) 'Single Molecule Assays of β -Galactosidase from Two Wild-type Strains of *E. coli*: Effects of Protease Inhibitors on Microheterogeneity and Different Relative Activities with Differing Substrates', *The Protein Journal*, 26(2), 95-105.
- Nichtl, A., Buchner, J., Jaenicke, R., Rudolph, R. and Scheibel, T. (1998) 'Folding and association of β -galactosidase', *Journal of Molecular Biology*, 282(5), 1083-1091.
- NIH (2002) *NIH Guidelines for Research involving recombinant DNA molecules*, Office of Biotechnology Activities.
- Nijboer, R. C. and Verdonschot, P. F. M. (2004) 'Variable selection for modelling effects of eutrophication on stream and river ecosystems', *Ecological Modelling*, 177(1-2), 17-39.
- Nikerel, I. E., Öner, E., Kirdar, B. and Yildirim, R. (2006) 'Optimization of medium composition for biomass production of recombinant *Escherichia coli* cells using response surface methodology', *Biochemical Engineering Journal*, 32(1), 1-6.
- Novosad, J. (1994) 'Phosphorus: Inorganic Chemistry' in King, B. R., ed. *Encyclopedia of Inorganic Chemistry*, 1st ed., Chichester: Wiley.
- OECD (1998) *Biotechnology for Clean Industrial Products and Processes: Towards Industrial Sustainability*, OECD Publishing.
- OECD (2010) *OECD Environmental Performance Reviews OECD Environmental Performance Reviews: Ireland 2010*, OECD Publishing.
- Ott, C. and Rechberger, H. (2012) 'The European phosphorus balance', *Resources, Conservation and Recycling*, 60(0), 159-172.
- Palomares, L., Estrada-Moncada, S. and Ramírez, O. (2004) 'Production of Recombinant Proteins' in Balbás, P. and Lorence, A., eds., *Recombinant Gene Expression*, Humana Press, 15-51.
- Paul, E., Laval, M. L. and Sperandio, M. (2001) 'Excess Sludge Production and Costs Due to Phosphorus Removal', *Environmental Technology*, 22(11), 1363-1371.
- Petty, K. J. (2001) 'Metal-Chelate Affinity Chromatography' in *Current Protocols in Protein Science*, John Wiley & Sons, Inc.
- Phillips, A. T. and Signs, M. W. (2001) 'Desalting, Concentration, and Buffer Exchange by Dialysis and Ultrafiltration' in *Current Protocols in Protein Science*, John Wiley & Sons, Inc.

- Pierzynski, G. M., Sims, J. T. and Vance, G. F. (2000) *Soils and environmental quality*, 2nd ed., Boca Raton: CRC Press.
- Pietrzykowski, M., Flanagan, W., Pizzi, V., Brown, A., Sinclair, A. and Monge, M. (2013) 'An environmental life cycle assessment comparison of single-use and conventional process technology for the production of monoclonal antibodies', *Journal of Cleaner Production*, 41(0), 150-162.
- Pissarra, P. d. N. (2004) *Recombinant DNA Proteins for the Biopharmaceutical Industry and the future of Escherichia coli*, *Business Briefing: Pharma Outsourcing*, Touch Briefings.
- Prescott, L. M., Harley, J. P. and Klein, D. A. (2002) *Microbiology*, 5th ed., Boston: McGraw Hill.
- Qian, Y., Huang, Z. and Yan, Z. (2007) 'Integrated Assessment of Environmental and Economic Performance of Chemical Products Using Analytic Hierarchy Process Approach', *Chinese Journal of Chemical Engineering*, 15(1), 81-87.
- Rader, R. A. (2013) 'FDA Biopharmaceutical Product Approvals and Trends in 2012', *BioProcess International*, 11(3), 18-27.
- Ramírez, E. C., Johnston, D. B., McAloon, A. J. and Singh, V. (2009) 'Enzymatic corn wet milling: engineering process and cost model', *Biotechnology for Biofuels*, 2(2).
- Ramzan, N., Degenkolbe, S. and Witt, W. (2008) 'Evaluating and improving environmental performance of HC's recovery system: A case study of distillation unit', *Chemical Engineering Journal*, 140(1-3), 201-213.
- Ransohoff, T. C. (2010) 'Biomanufacturing Capacity: The Bottleneck Moves Downstream', in *ISPE Annual Meeting*, Orlando, FL, 7th-10th Nov.
- Ratledge, C. A. and Kristiansen, B. A., eds. (2001) *Basic Biotechnology*, Cambridge: Cambridge University Press.
- Rebitzer, G., Ekvall, T., Frischknecht, R., Hunkeler, D., Norris, G., Rydberg, T., Schmidt, W. P., Suh, S., Weidema, B. P. and Pennington, D. W. (2004) 'Life cycle assessment: Part 1: Framework, goal and scope definition, inventory analysis, and applications', *Environment International*, 30(5), 701-720.
- Revuelta, J., Machetti, F. and Cicchi, S. (2011) 'Five-Membered Heterocycles: 1,3-Azoles' in Alvarez-Builla, J., Vaquero, J. J. and Barluenga, J., eds., *Modern Heterocyclic Chemistry*, Weinheim: Wiley, 809-924.
- Riedewald, F., Byrne, E. and Cronin, K. (2012) 'A stochastic model for performance analysis of pharmaceutical high purity water systems', *Simulation Modelling Practice and Theory*, 21(1), 26-38.
- Ring, M. and Huber, R. E. (1990) 'Multiple replacements establish the importance of tyrosine-503 in β -galactosidase (*Escherichia coli*)', *Archives of Biochemistry and Biophysics*, 283(2), 342-350.

- Roche Diagnostics (2010) 'Restriction Enzymes - Double digest table', [online], available: http://www.roche-applied-science.com/sis/cloning/index.jsp?id=mcct_0101011201 [accessed Dec2008].
- Roos, P. H. (1999) 'Ion Exchange Chromatography' in Kastner, M., ed. *Protein Liquid Chromatography*, Amsterdam: Elsevier, 3-88.
- Rosemarin, A., De Bruijne, G. and Caldwell, I. (2009) 'Peak Phosphorus: The next inconvenient truth', *The Broker*, August, 6-9.
- Roth, N. J. and Huber, R. E. (1996a) 'GLU-416 of β -Galactosidase (*Escherichia coli*) Is a MG²⁺Ligand and β -Galactosidases with Substitutions for GLU-416 Are Inactivated, Rather than Activated, by MG²⁺', *Biochemical and Biophysical Research Communications*, 219(1), 111-115.
- Roth, N. J. and Huber, R. E. (1996b) 'The β -Galactosidase (*Escherichia coli*) Reaction Is Partly Facilitated by Interactions of His-540 with the C6 Hydroxyl of Galactose', *Journal of Biological Chemistry*, 271(24), 14296-14301.
- Roth, N. J., Penner, R. M. and Huber, R. E. (2003) ' β -Galactosidases (*Escherichia coli*) with Double Substitutions Show That Tyr-503 Acts Independently of Glu-461 but Cooperatively with Glu-537', *Journal of Protein Chemistry*, 22(7), 663-668.
- Russell, D. L. (2006) 'Phosphorus Removal' in *Practical Wastewater Treatment*, Hoboken, N.J.: John Wiley & Sons, Inc., 191-202.
- Ryding, S.-O. and Rast, W. (1989) *The control of eutrophication of lakes and reservoirs, Man and the Biosphere*, Paris: Unesco - The Parthenon Publishing Group.
- Saaty, L. and Vargas, L. G. (2012) *Models, Methods, Concepts and Applications of the Analytic Hierarchy Process*, 2nd ed., New York: Springer.
- Sage, A. P. (2009) 'Systematic measurements' in Sage, A. P. and Rouse, W. B., eds., *Handbook of Systems Engineering and Management*, New Jersey: Wiley, 575-644.
- SAIC (2006) *Life Cycle Assessment: Principles and Practice*, Cincinnati: U.S. Environmental Protection Agency.
- Sambrook, J. and Russell, D. W. (2001) *Molecular Cloning. A laboratory Manual*, 3rd ed., New York: Cold Spring Harbor Laboratory Press.
- SCBD (2013) 'Article 2. Use of Terms', [online], available: <http://www.cbd.int/convention/articles/default.shtml?a=cbd-02> [accessed 04June2013].
- Schaechter, M. (2010) '*Escherichia coli*' in Schaechter, M., ed. *Desk Encyclopedia of Microbiology*, Oxford: Elsevier Science.
- Seppälä, J., Melanen, M., Jouttijärvi, T., Kauppi, L. and Leikola, N. (1998) 'Forest industry and the environment: a life cycle assessment study from Finland', *Resources, Conservation and Recycling*, 23(1-2), 87-105.

- Šeputienė, V., Daugelavičius, A., Sužiedėlis, K. and Sužiedėlienė, E. (2006) 'Acid response of exponentially growing *Escherichia coli* K-12', *Microbiological Research*, 161(1), 65-74.
- Shaked, S. and Jolliet, O. (2011) 'Global Life Cycle Impacts of Consumer Products' in Nriagu, J. O., ed. *Encyclopedia of Environmental Health*, Burlington: Elsevier, 1002-1014.
- Sharpley, A. (1999) 'Agricultural phosphorus, water quality, and poultry production: are they compatible?', *Poultry Science*, 78(5), 660-673.
- Shiloach, J. and Fass, R. (2005) 'Growing *E. coli* to high cell density--A historical perspective on method development', *Biotechnology Advances*, 23(5), 345-357.
- Shojaosadati, S. A., Kolaei, S. M. V., Babaeipour, V. and Mohammad, F. A. (2008) 'Recent advances in high cell density cultivation for production of recombinant protein', *Iranian Journal of Biotechnology*, 6(2), 63-84.
- Shu, L., Schneider, P., Jegatheesan, V. and Johnson, J. (2006) 'An economic evaluation of phosphorus recovery as struvite from digester supernatant', *Bioresource Technology*, 97(17), 2211-2216.
- Shuler, M. L. and Kargi, F. (2002) *Bioprocess Engineering*, 2nd ed., New Jersey: Prentice Hall PTR.
- Sigma-Aldrich (2013a) 'MSDS Search', [online], available: <http://www.sigmaaldrich.com/safety-center.html> [accessed 23Sep13].
- Sigma-Aldrich (2013b) 'Trypsin Inhibitors', [online], available: <http://www.sigmaaldrich.com/life-science/metabolomics/enzyme-explorer/analytical-enzymes/trypsin/trypsin-inhibitors.html#ref14> [accessed 27Aug2013].
- Sivashanmugam, A., Murray, V., Cui, C., Zhang, Y., Wang, J. and Li, Q. (2009) 'Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*', *Protein Science*, 18(5), 936-948.
- Soetaert, W. and Vandamme, E. J. (2010) *Industrial Biotechnology*, Weinheim: Wiley VCH Verlag GmbH & Co. KGaA.
- Soini, J., Ukkonen, K. and Neubauer, P. (2008) 'High cell density media for *Escherichia coli* are generally designed for aerobic cultivations - consequences for large-scale bioprocesses and shake flask cultures', *Microbial Cell Factories*, 7(1), 26-37.
- Sørensen, H. and Mortensen, K. (2005) 'Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*', *Microbial Cell Factories*, 4(1).
- Srivastava, P. and Mukherjee, K. J. (2005) 'Kinetic studies of recombinant human interferon-alpha (rhIFN- α) expression in transient state continuous cultures', *Biochemical Engineering Journal*, 26(1), 50-58.

- Steers, E., Jr., Cuatrecasas, P. and Pollard, H. B. (1971) 'The Purification of β -Galactosidase from *Escherichia coli* by Affinity Chromatography', *Journal of Biological Chemistry*, 246(1), 196-200.
- Stevenson, R. J., Bothwell, M. L., Lowe, R. L. and Thorp, J. H. (1996) *Algal Ecology: Freshwater Benthic Ecosystem*, San Diego: Academic Press.
- Stiborova, H., Kostal, J., Mulchandani, A. and Chen, W. (2003) 'One-step metal-affinity purification of histidine-tagged proteins by temperature-triggered precipitation', *Biotechnology and Bioengineering*, 82(5), 605-611.
- Struijs, J., Beusen, A., de Zwart, D. and Huijbregts, M. (2011) 'Characterization factors for inland water eutrophication at the damage level in life cycle impact assessment', *The International Journal of Life Cycle Assessment*, 16(1), 59-64.
- Suárez, D. C., Liria, C. W. and Kilikian, B. V. (1998) 'Effect of yeast extract on *Escherichia coli* growth and acetic acid production', *World Journal of Microbiology and Biotechnology*, 14(3), 331-335.
- Subramanian, G. (2012) *Biopharmaceutical Production Technology*, Weinheim: Wiley-VCH.
- Suh, S., Lenzen, M., Treloar, G. J., Hondo, H., Horvath, A., Huppes, G., Jolliet, O., Klann, U., Krewitt, W., Moriguchi, Y., Munksgaard, J. and Norris, G. (2003) 'System Boundary Selection in Life-Cycle Inventories Using Hybrid Approaches', *Environmental Science & Technology*, 38(3), 657-664.
- Taras, S. and Woinaroschy, A. (2012) 'An interactive multi-objective optimization framework for sustainable design of bioprocesses', *Computers & Chemical Engineering*, 43(0), 10-22.
- Tartoff, K. D. and Hobbs, C. A. (1987) 'Improved media for growing plasmid and cosmid clones', *Bethesda Research Laboratories Focus*, 9(2), 12-14.
- Tchobanoglous, G., Burton, F. L. and Stensel, H. D. (2003) *Wastewater Engineering: Treatment and Reuse*, 3rd ed., New York: McGraw-Hill Education.
- Tiessen, H. (2008) 'Phosphorus in the global environment' in White, P. J. and Hammond, J. P., eds., *The Ecophysiology of Plant-Phosphorus Interactions* Dordrecht: Springer Netherlands, 1-7.
- Tripathi, N. K., Babu, J. P., Shrivastva, A., Parida, M., Jana, A. M. and Rao, P. V. L. (2008) 'Production and characterization of recombinant dengue virus type 4 envelope domain III protein', *Journal of Biotechnology*, 134(3-4), 278-286.
- Tripathi, N. K., Sathyaseelan, K., Jana, A. M. and Rao, P. V. L. (2009a) 'High Yield Production of Heterologous Proteins with *Escherichia coli*', *Defence Science Journal*, 59(2), 137-146.
- Tripathi, N. K., Shrivastva, A., Biswal, K. C. and Lakshmana Rao, P. V. (2009b) 'Optimization of culture medium for production of recombinant dengue protein in *Escherichia coli*', *Industrial Biotechnology*, 5(3), 179-183.

- Ullmann, A. (2001) '*Escherichia coli* Lactose Operon' in *Encyclopedia of Life Sciences*, New York: John Wiley & Sons.
- Vaccari, D. A. (2009) 'Phosphorus: A Looming Crisis', *Scientific American*, 300(6), 54-59.
- Valsami-Jones, E. (2004) *Phosphorus in Environmental Technologies: Principles and Applications*, London: IWA Publishing.
- Van De Meent, D. and De Bruijn, J. H. M. (2007) 'Environmental Exposure Assessment' in van Leeuwen, C. J. and Vermeire, T. G., eds., *Risk Assessment of Chemicals: An Introduction*, Dordrecht: Springer, 159-194.
- Van der Vorst, G., Dewulf, J., Aelterman, W., De Witte, B. and Van Langenhove, H. (2009) 'Assessment of the Integral Resource Consumption of Individual Chemical Production Processes in a Multipurpose Pharmaceutical Production Plant: A Complex Task', *Industrial & Engineering Chemistry Research*, 48(11), 5344-5350.
- van Haandel, A. C. and van der Lubbe, J. G. M. (2012) 'Phosphorus removal' in *Handbook of Biological Wastewater Treatment: Design and Optimisation of Activated Sludge Systems*, London: IWA Publishing, 207-258.
- vanLoon, G. W. and Duffy, S. J. (2000) *Environmental Chemistry. A global perspective.*, Oxford: Oxford University Press.
- Vázquez, M. V. and Rodríguez, M. E. G. (2011) 'Computer simulation of microbial transglutaminase production from potato wastes', *Procedia Food Science*, 1(0), 736-741.
- Ventura, S. and Villaverde, A. (2006) 'Protein quality in bacterial inclusion bodies', *Trends in Biotechnology*, 24(4), 179-185.
- Viaplana, E., Feliu, J. X., Corchero, J. L. and Villaverde, A. (1997) 'Reversible activation of a cryptic cleavage site within *E. coli* β -galactosidase in β -galactosidase fusion proteins', *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, 1343(2), 221-226.
- Waites, M. J., Morgan, N. L., Rockey, J. S. and Higton, G. (2001) *Industrial Microbiology: An Introduction*, Oxford: Blackwell Science.
- Wallenfels, K. and Malhotra, O. M. (1960) ' β -Galactosidase' in Boyer, P. D., Lardy, H. and Myrbäck, K., eds., *The Enzymes*, 2nd ed., New York: Academic Press, 409-430.
- Wallenfels, K. and Weil, R. (1972) ' β -Galactosidase' in Boyer, P. D., ed. *The Enzymes*, 3rd ed., New York: Academic Press, 617-663.
- Walsh, G. (2000) 'Biopharmaceutical benchmarks', *Nature Biotechnology*, 18(8), 831-833.
- Walsh, G. (2002) *Proteins: Biochemistry and Biotechnology*, Chichester: Wiley.
- Walsh, G. (2003a) 'Biopharmaceutical benchmarks - 2003', *Nature Biotechnology*, 21(8), 865-870.

- Walsh, G. (2003b) *Biopharmaceuticals: biochemistry and biotechnology*, 2nd ed., Chichester: Wiley.
- Walsh, G. (2006) 'Biopharmaceutical benchmarks 2006', *Nature Biotechnology*, 24(7), 769-776.
- Walsh, G. (2007a) 'Biopharmaceuticals: Approval trends in 2006.', *Biopharm International*, 21(9), 40-48.
- Walsh, G. (2007b) *Pharmaceutical Biotechnology: Concepts and Applications*, Chichester: Wiley.
- Walsh, G., ed. (2009) *Post-translational Modification of Protein Biopharmaceuticals*, Weinheim: WILEY-VCH Verlag GmbH & Co. KGaA.
- Walsh, G. (2010a) 'Biopharmaceutical benchmarks 2010', *Nature Biotechnology*, 28(9), 917-924.
- Walsh, G. (2010b) 'Post-translational modifications of protein biopharmaceuticals', *Drug Discovery Today*, 15(17-18), 773-780.
- Walsh, G. (2012) 'New Biopharmaceuticals: A review of new biologic drug approvals over the years, featuring highlights from 2010 and 2011.', *Biopharm International*, 25(6), 34-38.
- Walter, J. (1999) 'Scale-up of Downstream Processing' in Kastner, M., ed. *Protein Liquid Chromatography*, Amsterdam: Elsevier, 765-783.
- Wang, Q., Vural Gürsel, I., Shang, M. and Hessel, V. (2013) 'Life cycle assessment for the direct synthesis of adipic acid in microreactors and benchmarking to the commercial process', *Chemical Engineering Journal*, 234(0), 300-311.
- Waugh, D. S. (2005) 'Making the most of affinity tags', *Trends in Biotechnology*, 23(6), 316-320.
- Wenzel, H., Hauschild, M. Z. and Alting, L. (2000) *Environmental Assessment of Products - Volume 1: Methodology, Tools and Case Studies in Product Development*, 3rd ed., Boston: Kluwer Academic Publishers.
- Wheater, P. C. and Cook, P. A. (2000) *Using Statistics to Understand the Environment, Routledge Introductions to Environment Series*, 1st ed., London: Routledge.
- Wheatley, R. W., Kappelhoff, J. C., Hahn, J. N., Dugdale, M. L., Dutkoski, M. J., Tamman, S. D., Fraser, M. E. and Huber, R. E. (2012) 'Substitution for Asn460 Cripples β -galactosidase (*Escherichia coli*) by increasing substrate affinity and decreasing transition state stability', *Archives of Biochemistry and Biophysics*, 521(1-2), 51-61.
- White, B. A., ed. (1993) *PCR Protocols: Current Methods and Applications*, Totowa, NJ: Humana Press Inc.
- Wilkinson, D. L. and Harrison, R. G. (1991) 'Predicting the Solubility of Recombinant Proteins in *Escherichia coli*', *Nature Biotechnology*, 9(5), 443-448.

- Williams, A. and Frasca, V. (2001) 'Ion-Exchange Chromatography' in *Current Protocols in Protein Science*, John Wiley & Sons, Inc., 8.2.1-8.2.30.
- Worsfold, P. J., Gimbert, L. J., Mankasingh, U., Omaka, O. N., Hanrahan, G., Gardolinski, P. C. F. C., Haygarth, P. M., Turner, B. L., Keith-Roach, M. J. and McKelvie, I. D. (2005) 'Sampling, sample treatment and quality assurance issues for the determination of phosphorus species in natural waters and soils', *Talanta*, 66(2), 273-293.
- Wu-Pong, S. and Rojanasakul, Y. (2008) *Biopharmaceutical drug design and development*, 2nd ed., Totowa: Humana Press, Springer.
- Yan, M.-J., Humphreys, J. and Holden, N. M. (2011) 'An evaluation of life cycle assessment of European milk production', *Journal of Environmental Management*, 92(3), 372-379.
- Yang, I. T., Wang, W.-C. and Yang, T.-I. (2012) 'Automatic repair of inconsistent pairwise weighting matrices in analytic hierarchy process', *Automation in Construction*, 22(0), 290-297.
- Yu, Q., Zhixian, H. and Zhiguo, Y. A. N. (2007) 'Integrated Assessment of Environmental and Economic Performance of Chemical Products Using Analytic Hierarchy Process Approach', *Chinese Journal of Chemical Engineering*, 15(1), 81-87.
- Yuan, Z., Pratt, S. and Batstone, D. J. (2012) 'Phosphorus recovery from wastewater through microbial processes', *Current Opinion in Biotechnology*, 23(6), 878-883.
- Zhang, J. and Greasham, R. (1999) 'Chemically defined media for commercial fermentations', *Applied Microbiology and Biotechnology*, 51(4), 407-421.
- Zhang, J., Reddy, J., Buckland, B. and Greasham, R. (2003) 'Toward consistent and productive complex media for industrial fermentations: Studies on yeast extract for a recombinant yeast fermentation process', *Biotechnology and Bioengineering*, 82(6), 640-652.
- Zhang, Y., Guo, Y.-j., Sun, S.-h., Yan, H.-l. and He, Y. (2004) 'Non-fusion expression in *Escherichia coli*, purification, and characterization of a novel Ca²⁺- and phospholipid-binding protein annexin B1', *Protein Expression and Purification*, 34(1), 68-74.
- Zhang, Z., Wu, X., Yang, X. and Zhu, Y. (2006) 'BEPAS—a life cycle building environmental performance assessment model', *Building and Environment*, 41(5), 669-675.
- Zhao, Y., Xi, B., Li, Y., Wang, M., Zhu, Z., Xia, X., Zhang, L.-y., Wang, L. and Luan, Z. (2012) 'Removal of phosphate from wastewater by using open gradient superconducting magnetic separation as pretreatment for high gradient superconducting magnetic separation', *Separation and Purification Technology*, 86(0), 255-261.
- Zhou, J., Chang, V. W. C. and Fane, A. G. (2011) 'Environmental life cycle assessment of reverse osmosis desalination: The influence of different life cycle impact assessment methods on the characterization results', *Desalination*, 283(0), 227-236.

Appendices

Appendix A: Protein and Gene sequences for β -galactosidase from *Escherichia coli* K12 substrain W3110 accession number AP_000996.1

Protein Sequence

10	20	30	40	50	60
MTMITDSLAV	VLQRRDWENP	GVTQLNRLAA	HPPFASWRNS	EEARTDRPSQ	QLRSLNGEWR
70	80	90	100	110	120
FAWFPAPEAV	PESWLECDLP	EADTVVVPSP	WQMHGYDAPI	YTNVTYPITV	NPPFVPTENP
130	140	150	160	170	180
TGCYSLTFNV	DESWLQEQGT	RIIFDGVNSA	FHLWCNGRWV	GYGQDSRLPS	EFDLSAFLRA
190	200	210	220	230	240
GENRLAVMVL	RWSDGSYLED	QDMWRMSGIF	RDVSLHKKPT	TQISDFHVAT	RFNDDFSRAV
250	260	270	280	290	300
LEAEVQMCGE	LRDYLRVTVS	LWQGETQVAS	GTAPFGGEII	DERGGYADRV	TLRLNVENPK
310	320	330	340	350	360
LWSAEIPNLY	RAVVELHTAD	GTLIEAEACD	VGFREVRIEN	GLLLLNGKPL	LIRGVNRHEH
370	380	390	400	410	420
HPLHGQVMDE	QTMVQDILLM	KQNNFNAVRC	SHYPNHPLWY	TLCDRYGLYV	VDEANIETHG
430	440	450	460	470	480
MVPMNRLTDD	PRWLPAMSER	VTRMVQRDRN	HPSVIIWSLG	NESGHGANHD	ALYRWIKSVD
490	500	510	520	530	540
PSRPVQYEGG	GADTTATDII	CPMYARVDED	QPFPAVPKWS	IKKWLSPGGE	TRPLILCEYA
550	560	570	580	590	600
HAMGNSLGGF	AKYWQAFRQY	PRLQGGFVWD	WVDQSLIKYD	ENGNPWSAYG	GDFGDTPNDR
610	620	630	640	650	660
QFCMNGLVFA	DRTPHPALTE	AKHQQQFFQF	RLSGQTIEVT	SEYLFHRSDN	ELLHWMVALD
670	680	690	700	710	720
GKPLASGEVP	LDVAPQGKQL	IELPELPQPE	SAGQLWLTVR	VVQPNATAWS	EAGHISAWQQ
730	740	750	760	770	780
WRLAENLSVT	LPAASHAIPH	LTTSEMDFCI	ELGNKRWQFN	RQSGFLSQMW	IGDKKQLLTP
790	800	810	820	830	840
LRDQFTRAPL	DNDIGVSEAT	RIDPNAWVER	WKAAGHYQAE	AALLQCTADT	LADAVLITTA
850	860	870	880	890	900
HAWQHQQKTL	FISRKTYRID	GSGQMAITVD	VEVASDTPHP	ARIGLNCQLA	QVAERVNWLG
910	920	930	940	950	960
LGPQENYPDR	LTAACFDRWD	LPLSDMYTPY	VFPSENGLRC	GTRELNYGPH	QWRGDFQFNI
970	980	990	1000	1010	1020
SRYSQQQLME	TSHRHLHAE	EGTWLNIDGF	HMGIGGDDSW	SPSVSAEFQL	SAGRYHYQLV

WCQK

Gene Sequence

1 ttatittga caccagacca actggaatg gtagcgaccg gcgctcagct ggaattccgc
61 cgatactgac gggctccagg agtcgtgcc accaatcccc atatggaaac cgtcgatatt
121 cagccatgtg ccttctccg cgtgcagcag atggcgatgg ctggttcca tcagttgctg
181 ttgactgtag cggctgatgt tgaactggaa gtcgccgcgc cactggtgtg ggccataatt
241 caattegcgc gtccccgagc gcagaccgtt ttcgctcggg aagacgtacg gggtatacat

301 gtctgacaat ggcagatccc agcgggtcaaa acaggcggca gtaaggcggg cgggatagtt
361 ttcttgccgc cctaatccga gccagtttac ccgctctgct acctgcgcca gctggcagtt
421 caggccaatc cgcgcccggat gcggtgtatc gctcgccact tcaacatcaa cggtaatgca
481 catttgacca ctaccatcaa tccggtaggt tttccggctg ataaataagg tttcccctg
541 atgctgccac gcgtgagcgg tcgtaatcag caccgcatca gcaagtgtat ctgccgtgca
601 ctgcaacaac gctgcttcgg cctggtaatg gcccgcggcc tccagcgtt cgaccaggc
661 gtaggggtca atgcgggtcg cttcacttac gccaatgtcg ttatccagcg gtgcacgggt
721 gaactgatcg cgcagcggcg tcagcagttg tttttatcg ccaatccaca tctgtgaaag
781 aaagcctgac tggcgggtaa attgccaacg cttattacc agctcgatgc aaaaatccat
841 ttcgctggg gtcagatgcg ggatggcgtg ggacgcggcg gggagcgtca cactgaggtt
901 ttcgccaga cgccactgct gccaggcgt gatgtgccc gcttctgacc atgcggtcgc
961 gttcgggtgc actacgcgta ctgtgagcca gagttgccc gcgctctcc gctgcggtg
1021 ttcaggcagt tcaatcaact gttacctg tggagcgaca tccagaggca cttaccgct
1081 tgccagcggc ttaccatcca gcgccaccat ccagtgcagg agctcgttat cgctatgacg
1141 gaacaggtat tcgctggca cttcgatggt ttgcccggat aaacggaact ggaaaaactg
1201 ctgctggtgt tttgcttccg tcagcgtg atgcggcgtg cggtcggcaa agaccagacc
1261 gttcatacag aactggcgat cgttcggcgt atgccaaaa tcaccgccgt aagccgacca
1321 cggggtgccc tttcatcat attaatcag cgactgatcc acccagtc cc agacgaagcc
1381 gccctgtaaa cggggatact gacgaaacgc ctgccagtat ttagcgaaac cgccaagact
1441 gttaccatc gcgtgggcgt attcgcaaag gatcagcggg cgcgtctctc caggtagcga
1501 aagccatfff ttgatggacc attcggcac agccgggaag ggctggttt catccacgcg
1561 cgcgtacatc gggcaataa tateggtggc cgtggtgtcg gctccgccgc cttcactg
1621 caccggcggg gaaggatcga cagattgat ccagcgatac agcgcgtcgt gattagcgc
1681 gtggcctgat tcatcccca gcgaccagat gatcacactc ggggtattac gategcgctg
1741 caccattcgc gttacgcgtt cgctcatcgc cggtagccag cgcggatcat cgtcagacg
1801 attcattgce accatgccgt gggtttcaat attggttca tccaccat acaggcgt
1861 gcggtcgcac agcgtgtacc acagcggatg gttcggataa tgcgaacagc gcacggcgtt
1921 aaagttgttc tgcttcatca gcaggatate ctgcaccatc gctgctcat ccatgacctg
1981 accatgcaga ggatgatgct cgtgacgggt aacgcctcga atcagcaacg gcttcccgtt
2041 cagcagcagc agaccatfff caatccgcac ctcgcggaaa ccgacatcgc aggttctgc
2101 ttaatcagc gtgccgtcgg cgggtgtcag ttcaaccacc gcacgataga gattcgggat
2161 ttcggcgtc cacagtttcg ggttttcgac gttcagacgt agtgtgacgc gatcggcata
2221 accaccagc tcatcgataa tttaccgcc gaaagcgcg gtgccgctgg cgacctgct

2281 ttcacctgc cataaagaaa ctgtaccg taggtagtca cgcaactgc cgcacatctg
2341 aactcagcc tccagtacag cgcggctgaa atcatcatta aagcgagtgg caacatggaa
2401 atcgtgatt tgtgtagtcg gtttatgcag caacgagacg tcacggaaaa tgccgctcat
2461 ccgccacata tectgatctt ccagataact gccgtcactc cagegcagca ccatcaccgc
2521 gaggcggttt tctccggcgc gtaaaaatgc gtcaggta aattcagacg gcaaacgact
2581 gtctggccg taaccgacc agcgccggtt gcaccacaga tgaaacgcc agttaacgcc
2641 atcaaaaata attcgctctt ggccttctg tagccagctt tcatcaacat taaatgtgag
2701 cgagtaacaa cccgtcgat tctccgtggg aacaaacggc ggattgaccg taatgggata
2761 ggtcacgttg gtgtagatgg gcgcatcgt accgtgcatc tgccagttg aggggacgac
2821 gacagtatcg gcctcaggaa gatcgcactc cagccagctt tccggcaccg cttctggtgc
2881 cggaaaccag gcaaagcgc attcgccatt caggctgccc aactgttggg aaggcgatc
2941 ggtgcgggcc tcttcgctat tacgccagct ggcgaaaggg ggatgtgctg caaggcgatt
3001 aagttgggta acgccagggt tttccagtc acgacgttg aaaacgacgg ccagtgaatc
3061 cgtaatcatg gtcat

Appendix B: Primer design for *lacZ* amplification from *E. coli* W3110

β -Galactosidase is on the negative strand (362,455-365,529).

Reverse Primer

5' Hind III

3'

GGAAGCTTTTATTTTGGACACCAGACCAACTGG (T_m=58 °C)

5'

363450

|

CAGACATGGCCTGCCCGTTATTA**TTA**TTTTGACACCAGACCAACTGGTAATGGTA
GCGACCGGCGCTCAGCTGGAATTCCGCCGATACTGACGGGCTCCAGGAGTCGTCGCC
ACCAATCCCCATATGGAAACCGTCGATATTCAGCCATGTGCCTTCTCCGCGTGCAG
CAGATGGCGATGGCTGGTTTCCATCAGTTGCTGTTGACTGTAGCGGCTGATGTTGAA
CTGGAAGTCGCCGCGCCACTGGTGTGGGCCATAATTCAATTCGCGCGTCCCGCAGCG
CAGACCGTTTTCGCTCGGGAAGACGTACGGGGTATACATGTCTGACAATGGCAGATC
CCAGCGGTCAAACAGGCGGCAGTAAGGCGGTCTGGGATAGTTTTCTTGCGGCCCTA
ATCCGAGCCAGTTTACCGCTCTGCTACCTGCGCCAGCTGGCAGTTCAGGCCAATCC
GCGCCGGATGCGGTGTATCGCTCGCCACTTCAACATCAACGGTAATCGCCATTTGAC
CACTACCATCAATCCGGTAGGTTTTCCGGCTGATAAATAAGGTTTTCCCTGATGCTG
CCACGCGTGAGCGGTGTAATCAGCACCGCATCAGCAAGTGTATCTGCCGTGCACTG
CAACAACGCTGCTTCGGCCTGGTAATGGCCCGCCGCCTTCCAGCGTTCGACCCAGGC
GTTAGGGTCAATGCGGGTCGTTCACTTACGCCAATGTCGTTATCCAGCGGTGCACG
GGTGAAGTATCGCGCAGCGGCGTCAGCAGTTGTTTTTATCGCCAATCCACATCTG
TGAAAGAAAGCCTGACTGGCGGTTAAATTGCCAACGCTTATTACCCAGCTCGATGCA
AAAATCCATTTTCGCTGGTGGTCAGATGCGGGATGGCGTGGGACGCGGCGGGGAGCG
TCACACTGAGGTTTTCCGCCAGACGCCACTGCTGCCAGGCGCTGATGTGCCCGGCTT
CTGACCATGCGGTGCGGTTTCGGTTGCACTACGCGTACTGTGAGCCAGAGTTGCCCGG
CGCTCTCCGGCTGCGGTAGTTCAGGCAGTTCAATCAACTGTTTACCTTGTGGAGCGA
CATCCAGAGGCACTTCACCGCTTGCCAGCGGCTTACCATCCAGCGCCACCATCCAGT
GCAGGAGCTCGTTATCGCTATGACGGAACAGGTATTCGCTGGTCACTTCGATGGTTT
GCCCCGATAAACGGAAGTGGAAAACTGCTGCTGGTGTGTTTTGCTTCGTCAGCGCTG
GATGCGGCGTGCGGTGCGCAAAGACCAGACCGTTCATACAGAACTGGCGATCGTTC
GGCGTATCGCCAAAATCACCGCCGTAAGCCGACCACGGGTTGCCGTTTTTCATCATAT
TTAATCAGCGACTGATCCACCCAGTCCCAGACGAAGCCGCCCTGTAAACGGGGATAC
TGACGAAACGCCTGCCAGTATTTAGCGAAACCGCCAAGACTGTTACCCATCGCGTGG
GCGTATTCGCAAAGGATCAGCGGGCGCGTCTCTCCAGGTAGCGAAAGCCATTTTTTG
ATGGACCATTTTCGGCACAGCCGGAAGGGCTGGTCTTCATCCACGCGCGGTACATC

GGGCAAATAATATCGGTGGCCGTGGTGTCGGCTCCGCCGCCTTCATACTGCACCGGG
 CGGGAAGGATCGACAGATTTGATCCAGCGATACAGCGCGTCGTGATTAGCGCCGTG
 GCCTGATTCATTCCCAGCGACCAGATGATCACACTCGGGTGATTACGATCGCGCTG
 CACCATTGCGGTTACGCGTTCGCTCATCGCCGGTAGCCAGCGCGGATCATCGGTCAG
 ACGATTCAATTGGCACCATGCCGTGGGTTTCAATATTGGCTTCATCCACCACATACAG
 GCCGTAGCGGTTCGCACAGCGTGTACCACAGCGGATGGTTTCGGATAATGCGAACAGC
 GCACGGCGTTAAAGTTGTTCTGCTTCATCAGCAGGATATCCTGCACCATCGTCTGCTC
 ATCCATGACCTGACCATGCAGAGGATGATGCTCGTGACGGTTAACGCCTCGAATCAG
 CAACGGCTTGCCGTTACGACAGCAGCAGACCATTTTCAATCCGCACCTCGCGGAAACC
 GACATCGCAGGCTTCTGCTTCAATCAGCGTGCCGTCGGCGGTGTGCAGTTCAACCAC
 CGCACGATAGAGATTCGGGATTTTCGGCGCTCCACAGTTTCGGGTTTTTCGACGTTACG
 ACGTAGTGTGACGCGATCGGCATAACCACCACGCTCATCGATAATTCACCGCCGAA
 AGGCGCGGTGCCGCTGGCGACCTGCGTTTTACCCTGCCATAAAGAAACTGTTACCCG
 TAGGTAGTCACGCAACTCGCCGCACATCTGAACTTCAGCCTCCAGTACAGCGCGGCT
 GAAATCATCATTAAAGCGAGTGGCAACATGGAAATCGCTGATTTGTGTAGTCGGTTT
 ATGCAGCAACGAGACGTCACGGAAAATGCCGCTCATCCGCCACATATCCTGATCTTC
 CAGATAACTGCCGTCCTCAGCGCAGCACCATCACCGCGAGGCGGTTTTCTCCGGC
 GCGTAAAAATGCGCTCAGGTCAAATTCAGACGGCAAACGACTGTCCTGGCCGTAAC
 CGACCCAGCGCCCGTTGCACCACAGATGAAACGCCGAGTTAACGCCATCAAAAATA
 ATTCGCGTCTGGCCTTCCTGTAGCCAGCTTTCATCAACATTAAATGTGAGCGAGTAA
 CAACCCGTCGGATTCTCCGTGGGAACAAACGGCGGATTGACCGTAATGGGATAGGT
 CACGTTGGTGTAGATGGGCGCATCGTAACCGTGTCATCTGCCAGTTTGAGGGGACGAC
 GACAGTATCGGCCTCAGGAAGATCGCACTCCAGCCAGCTTTCGGGCACCGCTTCTGG
 TGCCGAAACCAGGCAAAGCGCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAG
 GGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTG
 CAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCAGTCACGACGTTGTAACGAA
 CGGCCAGTGAATCCGTAATCATGGT**CAT**AGCTGTTTCCTGTGT

|
365530 3'

GTCACCTAGGCATTAGTACCAGTACCTAGGTAG (T_m= 58 °C)
 3' Bam H1 5'

Forward Primer

Grey: *lacZ* gene

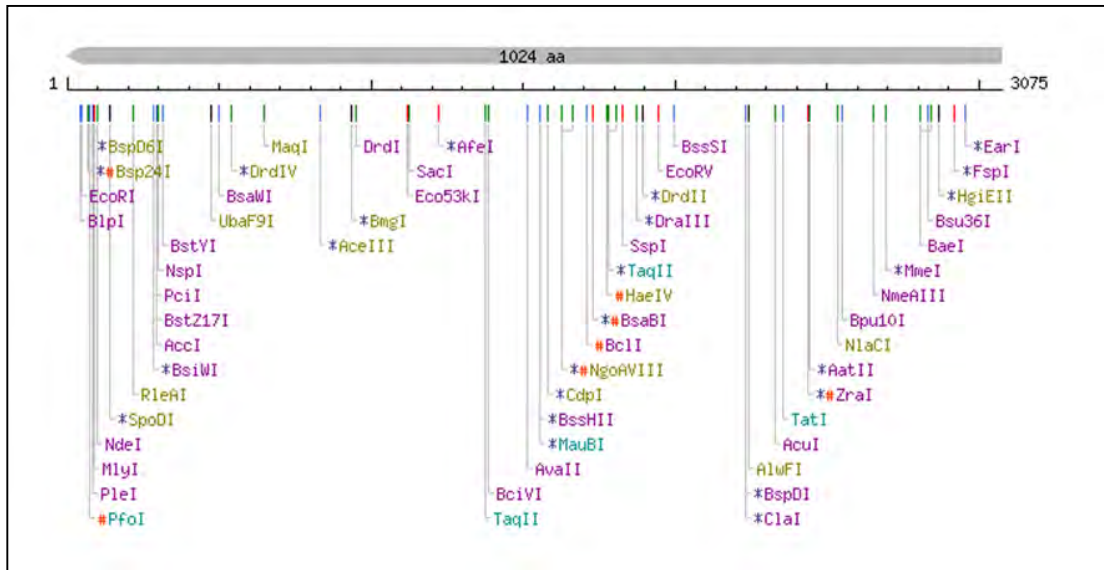
Yellow: overhang; used to adjust T_m

Red: Start/Stop codon

Green: Sequence of restriction enzyme

Blue: Annealing to gene sequence

Appendix C: Restriction digest of *lacZ* from *E. coli* W3110 using NEBCutter



Cleavage code	Enzyme name code
▼ blunt end cut	Available from NEB
▲ 5' extension	Has other supplier
▲ 3' extension	Not commercially available
▼ cuts 1 strand	*: cleavage affected by CpG meth.
	#: cleavage affected by other meth.
	(enz.name): ambiguous site

Appendix D: ClustalW alignment of recombinant *lacZ*-HT nucleotide sequence from MWG analysis and published gene sequence of *lacZ* from W3110 (Accession number AP_000996.1)

Nucleotide Sequence alignment of W3110_lacZ and lacZ-HT

SeqA Name	Len(nt)	SeqB Name	Len(nt)	Score
1 W3110_lacZ	3075	2 lacZ-HT	3075	99
W3110_lacZ	ATGACCATGATTACGGATTCACTGGCCGTCGTTTTTACAACGTCGTGACTGGGAAAACCT	lacZ-HT	ATGACCATGATTACGGATTCACTGGCCGTCGTTTTTACAACGTCGTGACTGGGAAAACCT	60
W3110_lacZ	GGCGTTACCCAACCTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGC	lacZ-HT	GGCGTTACCCAACCTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGC	120
W3110_lacZ	GAAAGAGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGC	lacZ-HT	GAAAGAGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGC	180
W3110_lacZ	TTTGCCCTGGTTTTCCGGCACCAGAAGCGGTGCCGAAAAGCTGGCTGGAGTGCATCTTCCT	lacZ-HT	TTTGCCCTGGTTTTCCGGCACCAGAAGCGGTGCCGAAAAGCTGGCTGGAGTGCATCTTCCT	240
W3110_lacZ	GAGGCCGATACTGTCGTCGCCCTCAAACGTCAGATGCACGGTTACGATGCGCCCATC	lacZ-HT	GAGGCCGATACTGTCGTCGCCCTCAAACGTCAGATGCACGGTTACGATGCGCCCATC	300
W3110_lacZ	TACACCAACGTGACCTATCCCATACGGTCAATCCGCCGTTTGTGCCACGGAGAATCCG	lacZ-HT	TACACCAACGTGACCTATCCCATACGGTCAATCCGCCGTTTGTGCCACGGAGAATCCG	360
W3110_lacZ	ACGGGTTGTTACTCGCTCACATTTAATGTTGATGAAAGCTGGCTACAGGAAGGCCAGACG	lacZ-HT	ACGGGTTGTTACTCGCTCACATTTAATGTTGATGAAAGCTGGCTACAGGAAGGCCAGACG	420
W3110_lacZ	CGAATTATTTTTGATGGCGTTAACTCGGCGTTTCATCTGTGGTGCAACGGGCGCTGGGTC	lacZ-HT	CGAATTATTTTTGATGGCGTTAACTCGGCGTTTCATCTGTGGTGCAACGGGCGCTGGGTC	480
W3110_lacZ	GGTTACGGCCAGGACAGTCGTTTGCCTGCTGAATTTGACCTGAGCGCATTTTTACGCGCC	lacZ-HT	GGTTACGGCCAGGACAGTCGTTTGCCTGCTGAATTTGACCTGAGCGCATTTTTACGCGCC	540
W3110_lacZ	GGAGAAAACCGCTCGCGGTGATGGTGTGCGCTGGAGTGACGGCAGTTATCTGGAAGAT	lacZ-HT	GGAGAAAACCGCTCGCGGTGATGGTGTGCGCTGGAGTGACGGCAGTTATCTGGAAGAT	600
W3110_lacZ	CAGGATATGTGGCGGATGAGCGGCATTTCCGTGACGTCTCGTTGCTGCATAAACCGACT	lacZ-HT	CAGGATATGTGGCGGATGAGCGGCATTTCCGTGACGTCTCGTTGCTGCATAAACCGACT	660
W3110_lacZ	ACACAAATCAGCGATTTCCATGTTGCCACTCGCTTTAATGATGATTTACGCCGCGCTGTA	lacZ-HT	ACACAAATCAGCGATTTCCATGTTGCCACTCGCTTTAATGATGATTTACGCCGCGCTGTA	720
W3110_lacZ	CTGGAGGCTGAAAGTTTCAGATGTGCGGCGAGTTGCGTGACTACCTACGGGTAACAGTTTCT	lacZ-HT	CTGGAGGCTGAAAGTTTCAGATGTGCGGCGAGTTGCGTGACTACCTACGGGTAACAGTTTCT	780
W3110_lacZ	TTATGGCAGGGTGAAACGCAGGTGCCAGCGGCACCGCGCCTTTCGGCGGTGAAATTATC	lacZ-HT	TTATGGCAGGGTGAAACGCAGGTGCCAGCGGCACCGCGCCTTTCGGCGGTGAAATTATC	840
W3110_lacZ	GATGAGCGTGGTGGTTATGCCGATCGCGTCACACTACGTCTGAACGTCGAAAACCCGAAA	lacZ-HT	GATGAGCGTGGTGGTTATGCCGATCGCGTCACACTACGTCTGAACGTCGAAAACCCGAAA	900

lacZ-HT GATGAGCGTGGTGGTTATGCCGATCGCGTCACACTACGTCTGAACGTCGAAAACCCGAAA 900

W3110_lacZ CTGTGGAGCGCCGAAATCCCGAATCTCTATCGTGCGGTGGTTGAACTGCACACCCGCCGAC 960
lacZ-HT CTGTGGAGCGCCGAAATCCCGAATCTCTATCGTGCGGTGGTTGAACTGCACACCCGCCGAC 960

W3110_lacZ GGCACGCTGATTGAAGCAGAAGCCTGCGATGTCGGTTTCCGCGAGGTGCGGATTGAAAAAT 1020
lacZ-HT GGCACGCTGATTGAAGCAGAAGCCTGCGATGTCGGTTTCCGCGAGGTGCGGATTGAAAAAT 1020

W3110_lacZ GGTCTGCTGCTGCTGAACGGCAAGCCGTGCTGATTGAGGCGTTAACCGTCACGAGCAT 1080
lacZ-HT GGTCTGCTGCTGCTGAACGGCAAGCCGTGCTGATTGAGGCGTTAACCGTCACGAGCAT 1080

W3110_lacZ CATCCTCTGCATGGTCAGGTCATGGATGAGCAGACGATGGTGCAGGATATCCTGCTGATG 1140
lacZ-HT CATCCTCTGCATGGTCAGGTCATGGATGAGCAGACGATGGTGCAGGATATCCTGCTGATG 1140

W3110_lacZ AAGCAGAACAACTTTAACGCCGTGCGCTGTTTCGCATTATCCGAACCATCCGCTGTGGTAC 1200
lacZ-HT AAGCAGAACAACTTTAACGCCGTGCGCTGTTTCGCATTATCCGAACCATCCGCTGTGGTAC 1200

W3110_lacZ ACGCTGTGCGACCGCTACGGCCTGTATGTGGTGGATGAAGCCAATATTGAAACCCACGGC 1260
lacZ-HT ACGCTGTGCGACCGCTACGGCCTGTATGTGGTGGATGAAGCCAATATTGAAACCCACGGC 1260

W3110_lacZ ATGGTGCCAATGAATCGTCTGACCGATGATCCGCGCTGGCTACCGGCATGAGCGAACGC 1320
lacZ-HT ATGGTGCCAATGAATCGTCTGACCGATGATCCGCGCTGGCTACCGGCATGAGCGAACGC 1320

W3110_lacZ GTAACGCGAATGGTGCAGCGCGATCGTAATCACCCGAGTGTGATCATCTGGTCGCTGGGG 1380
lacZ-HT GTAACGCGAATGGTGCAGCGCGATCGTAATCACCCGAGTGTGATCATCTGGTCGCTGGGG 1380

W3110_lacZ AATGAATCAGGCCACGGCGCTAATCACGACGCGCTGTATCGCTGGATCAAATCTGTTCGAT 1440
lacZ-HT AATGAATCAGGCCACGGCGCTAATCACGACGCGCTGTATCGCTGGATCAAATCTGTTCGAT 1440

W3110_lacZ CCTTCCCGCCCGTGCAGTATGAAGGCGCGGAGCCGACACCACGGCCACCGATATTATT 1500
lacZ-HT CCTTCCCGCCCGTGCAGTATGAAGGCGCGGAGCCGACACCACGGCCACCGATATTATT 1500

W3110_lacZ TGCCCGATGTACGCGCGCTGGATGAAGACCAGCCCTTCCCGGCTGTGCCGAAATGGTCC 1560
lacZ-HT TGCCCGATGTACGCGCGCTGGATGAAGACCAGCCCTTCCCGGCTGTGCCGAAATGGTCC 1560

W3110_lacZ ATCAAAAAATGGCTTTCGCTACCTGGAGAGACGCGCCCGCTGATCCTTTGCGAATACGCC 1620
lacZ-HT ATCAAAAAATGGCTTTCGCTACCTGGAGAGACGCGCCCGCTGATCCTTTGCGAATACGCC 1620

W3110_lacZ CACGCGATGGGTAACAGTCTTGCGGTTTCGCTAAATACTGGCAGGCGTTTCGTCAGTAT 1680
lacZ-HT CACGCGATGGGTAACAGTCTTGCGGTTTCGCTAAATACTGGCAGGCGTTTCGTCAGTAT 1680

W3110_lacZ CCCCGTTTACAGGGCGGCTTCGCTCTGGGACTGGGTGGATCAGTCGCTGATTAATATGAT 1740
lacZ-HT CCCCGTTTACAGGGCGGCTTCGCTCTGGGACTGGGTGGATCAGTCGCTGATTAATATGAT 1740

W3110_lacZ GAAAACGGCAACCCGTGGTFCGGCTTACGGCGGTGATTTTGGCGATAACGCCAAGATCGC 1800
lacZ-HT GAAAACGGCAACCCGTGGTFCGGCTTACGGCGGTGATTTTGGCGATAACGCCAAGATCGC 1800

W3110_lacZ CAGTTCGTATGAACGGTCTGGTCTTTGCCGACCGCACCCGCATCCAGCGCTGACGGAA 1860
lacZ-HT CAGTTCGTATGAACGGTCTGGTCTTTGCCGACCGCACCCGCATCCAGCGCTGACGGAA 1860

W3110_lacZ GCAAAACACCAGCAGCAGTTTTTCCAGTTCGGTTTATCCGGGCAAACCATCGAAGTGACC 1920
lacZ-HT GCAAAACACCAGCAGCAGTTTTTCCAGTTCGGTTTATCCGGGCAAACCATCGAAGTGACC 1920

W3110_lacZ AGCGAATACCTGTTCCGTATAGCGATAACGAGCTCCTGCACTGGATGGTGGCGCTGGAT 1980
lacZ-HT AGCGAATACCTGTTCCGTATAGCGATAACGAGCTCCTGCACTGGATGGTGGCGCTGGAT 1980

W3110_lacZ GGTAAGCCGCTGGCAAGCGGTGAAGTGCTCTGGATGTCGCTCCACAAGGTAACAGTTG 2040
lacZ-HT GGTAAGCCGCTGGCAAGCGGTGAAGTGCTCTGGATGTCGCTCCACAAGGTAACAGTTG 2040

W3110_lacZ ATTGAAGTGCCTGAACTACCGCAGCCGGAGAGCGCCGGGCAACTCTGGCTCACAGTACGC 2100
lacZ-HT ATTGAAGTGCCTGAACTACCGCAGCCGGAGAGCGCCGGGCAACTCTGGCTCACAGTACGC 2100

W3110_lacZ GTAGTGAACCGAACCGCACCAGCATGGTCAGAAAGCCGGGCACATCAGCGCTGGCAGCAG 2160
lacZ-HT GTAGTGAACCGAACCGCACCAGCATGGTCAGAAAGCCGGGCACATCAGCGCTGGCAGCAG 2160

W3110_lacZ TGGCGTCTGGCGGAAAACCTCAGTGTGACGCTCCCGCCGCGTCCCACGCCATCCCGCAT 2220
lacZ-HT TGGCGTCTGGCGGAAAACCTCAGTGTGACGCTCCCGCCGCGTCCCACGCCATCCCGCAT 2220

W3110_lacZ CTGACCACCAGCGAAATGGATTTTTCATCGAGCTGGGTAATAAGCGTTGGCAATTAAAC 2280
lacZ-HT CTGACCACCAGCGAAATGGATTTTTCATCGAGCTGGGTAATAAGCGTTGGCAATTAAAC 2280

W3110_lacZ CGCCAGTCAGGCTTTCTTTTCACAGATGTGGATTGGCGATAAAAAACAACCTGCTGACGCCG 2340
lacZ-HT CGCCAGTCAGGCTTTCTTTTCACAGATGTGGATTGGCGATAAAAAACAACCTGCTGACGCCG 2340

W3110_lacZ CTGCGGATCAGTTACCCGTGCACCGCTGGATAACGACATTGGCGTAAGTGAAGCGACC 2400
lacZ-HT CTGCGGATCAGTTACCCGTGCACCGCTGGATAACGACATTGGCGTAAGTGAAGCGACC 2400

W3110_lacZ CGCATTGACCCTAACGCCTGGGTGCAACGCTGGAAGGCGCGGGCCATTACCAGGCCGAA 2460
lacZ-HT CGCATTGACCCTAACGCCTGGGTGCAACGCTGGAAGGCGCGGGCCATTACCAGGCCGAA 2460

W3110_lacZ GCAGCGTTGTTGAGTGCACGGCAGATACACTTGCTGATGCGGTGCTGATTACGACCGCT 2520
lacZ-HT GCAGCGTTGTTGAGTGCACGGCAGATACACTTGCTGATGCGGTGCTGATTACGACCGCT 2520

W3110_lacZ CACGCGTGGCAGCATCAGGGGAAAACCTTATTTATCAGCCGGAAAACCTACCGGATTGAT 2580
lacZ-HT CACGCGTGGCAGCATCAGGGGAAAACCTTATTTATCAGCCGGAAAACCTACCGGATTGAT 2580

W3110_lacZ GGTAGTGGTCAAATGGCGATTACCGTTGATGTTGAAGTGGCGAGCGATACACCGCATCCG 2640
lacZ-HT GGTAGTGGTCAAATGGCGATTACCGTTGATGTTGAAGTGGCGAGCGATACACCGCATCCG 2640

W3110_lacZ GCGCGGATTGGCTGAACTGCCAGCTGGCGCAGGTAGCAGAGCGGGTAAACTGGCTCGGA 2700
lacZ-HT GCGCGGATTGGCTGAACTGCCAGCTGGCGCAGGTAGCAGAGCGGGTAAACTGGCTCGGA 2700

W3110_lacZ TTAGGGCCGCAAGAAAACCTATCCCGACCGCCTTACTGCCGCCTGTTTTGACCGCTGGGAT 2760
lacZ-HT TTAGGGCCGCAAGAAAACCTATCCCGACCGCCTTACTGCCGCCTGTTTTGACCGCTGGGAT 2760

W3110_lacZ CTGCCATTGTGACAGATGTATACCCCGTACGCTTCCCGAGCGAAAACGGTCTGCGCTGC 2820
lacZ-HT CTGCCATTGTGACAGATGTATACCCCGTACGCTTCCCGAGCGAAAACGGTCTGCGCTGC 2820

W3110_lacZ GGGACGCGCAATTGAATTATGGCCACACCAGTGGCGCGGCGACTTCCAGTTCAACATC 2880
lacZ-HT GGGACGCGCAATTGAATTATGGCCACACCAGTGGCGCGGCGACTTCCAGTTCAACATC 2880

W3110_lacZ AGCCGCTACAGTCAACAGCAACTGATGGAAACCAGCCATCGCCATCTGCTGCACGCGGAA 2940
lacZ-HT AGCCGCTACAGTCAACAGCAACTGATGGAAACCAGCCATCGCCATCTGCTGCACGCGGAA 2940

W3110_lacZ GAAGGCACATGGCTGAATATCGACGGTTTCCATATGGGGATTGGTGGCGGACACTCCTGG 3000
lacZ-HT GAAGGCACATGGCTGAATATCGACGGTTTCCATATGGGGATTGGTGGCGGACACTCCTGG 3000

W3110_lacZ AGCCCGTCAGTATCGGCGGAAATCCAGCTGAGCGCCGGTCGCTACCATTACCAGTTGGTC 3060
lacZ-HT AGCCCGTCAGTATCGGCGGAAATCCAGCTGAGCGCCGGTCGCTACCATTACCAGTTGGTC 3060

W3110_lacZ TGGTGTCAAAAATAA 3075
lacZ-HT TGGTGTCAAAAATAA 3075

Appendix E: Amino acid alignment of β -galactosidase from *E. coli* W3110 (accession number AP_000996.1) and recombinant β -galactosidase

SeqA Name	Len(aa)	SeqB Name	Len(aa)	Score
1 W3110_lacZ	1076	2 lacZ-HT	1076	100
W3110_lacZ	METTMETITDSLAVVLQRRDWDENPGVTQLNRLAAHPPFASWRNSEEARTDRPSQQLRSLN	60		
lacZ-HT	METTMETITDSLAVVLQRRDWDENPGVTQLNRLAAHPPFASWRNSEEARTDRPSQQLRSLN	60		

W3110_lacZ	GEWRFAPFPAPPEAVPESWLECDLPEADTVVPSNWQMETHGYDAPITYTNVTYPITVNPFF	120		
lacZ-HT	GEWRFAPFPAPPEAVPESWLECDLPEADTVVPSNWQMETHGYDAPITYTNVTYPITVNPFF	120		

W3110_lacZ	VPTENPTGCYSLTFNVDSEWLQEGQTRIFDGVNSAFHLWCNGRWVGYGQDSRLPSEFDL	180		
lacZ-HT	VPTENPTGCYSLTFNVDSEWLQEGQTRIFDGVNSAFHLWCNGRWVGYGQDSRLPSEFDL	180		

W3110_lacZ	SAFLRAGENRLAVMETVLRWSDGSYLEQDMETWRMETSIGIFRDVSLHKKPTTQISDFHV	240		
lacZ-HT	SAFLRAGENRLAVMETVLRWSDGSYLEQDMETWRMETSIGIFRDVSLHKKPTTQISDFHV	240		

W3110_lacZ	ATRFNDDFSRAVLEAEVQMETCGELRDYLRVTVSLWQGETQVASGTAPFGGEIIDERGGY	300		
lacZ-HT	ATRFNDDFSRAVLEAEVQMETCGELRDYLRVTVSLWQGETQVASGTAPFGGEIIDERGGY	300		

W3110_lacZ	ADRVTLRNLNVENPKLWSAEIPNLYRAVVELHTADGTLIEAEACDVGFEVRIENGLLLL	360		
lacZ-HT	ADRVTLRNLNVENPKLWSAEIPNLYRAVVELHTADGTLIEAEACDVGFEVRIENGLLLL	360		

W3110_lacZ	GKPLLIRGVNRHEHHPLHGQVMETDEQTMETVQDILLMETKQNNFNAVRCSHYPNHLWY	420		
lacZ-HT	GKPLLIRGVNRHEHHPLHGQVMETDEQTMETVQDILLMETKQNNFNAVRCSHYPNHLWY	420		

W3110_lacZ	TLCDRYGLYVVDEANIETHGMETVPMETNRLTDDPRWLPAMETSERVTRMETVQRDRNHP	480		
lacZ-HT	TLCDRYGLYVVDEANIETHGMETVPMETNRLTDDPRWLPAMETSERVTRMETVQRDRNHP	480		

W3110_lacZ	SVIIWSLGNESGHGANHDALYRWIKSVDPSPRVQYEGGADTTATDIICPMETYARVDED	540		
lacZ-HT	SVIIWSLGNESGHGANHDALYRWIKSVDPSPRVQYEGGADTTATDIICPMETYARVDED	540		

W3110_lacZ	QPFPAVPKWSIKKWLSPGETRPLILCEYAHAMETGNSLGGFAKYWQAFRQYPRLQGGFV	600		
lacZ-HT	QPFPAVPKWSIKKWLSPGETRPLILCEYAHAMETGNSLGGFAKYWQAFRQYPRLQGGFV	600		

W3110_lacZ	WDWVDQSLIKYDENGPNWSAYGGDFGDTPNDRQFCMETNGLVFAADRTPHPALTEAKHQQQ	660		
lacZ-HT	WDWVDQSLIKYDENGPNWSAYGGDFGDTPNDRQFCMETNGLVFAADRTPHPALTEAKHQQQ	660		

W3110_lacZ	FFQFRLSGQTIEVTSEYLFHRSDNELLHWMETVALDGKPLASGEVPLDVAPQKGQLIELP	720		
lacZ-HT	FFQFRLSGQTIEVTSEYLFHRSDNELLHWMETVALDGKPLASGEVPLDVAPQKGQLIELP	720		

W3110_lacZ	ELPQPESAGQLWLTVRVVPQPNATAWSEAGHISAWQQWRLAENLSVTLPAASHAIPLHTTS	780		
lacZ-HT	ELPQPESAGQLWLTVRVVPQPNATAWSEAGHISAWQQWRLAENLSVTLPAASHAIPLHTTS	780		

W3110_lacZ	EMETDFCIELGNKRWQFNRSQSGFLSQMETWIGDKKQLLTPLRDQFTRAPLDNDIGVSEAT	840		
lacZ-HT	EMETDFCIELGNKRWQFNRSQSGFLSQMETWIGDKKQLLTPLRDQFTRAPLDNDIGVSEAT	840		

W3110_lacZ	RIDPNAWVERWKAAGHYQAEAAALQCTADTLADAVLITTAHAWQHGGKTLFISRKYRID	900		
lacZ-HT	RIDPNAWVERWKAAGHYQAEAAALQCTADTLADAVLITTAHAWQHGGKTLFISRKYRID	900		

```
W3110_lacZ      GSGQMETAITVDVEVASDTPHPARIGLNCQLAQAERVNWLGLGPQENYPDRLTAACFDR 960
lacZ-HT        GSGQMETAITVDVEVASDTPHPARIGLNCQLAQAERVNWLGLGPQENYPDRLTAACFDR 960
                *****

W3110_lacZ      WDLPLSDMETYTPYVFPSENGLRGCTRELNYGPHQWRGDFQFNISRYSQQLMETETSHR 1020
lacZ-HT        WDLPLSDMETYTPYVFPSENGLRGCTRELNYGPHQWRGDFQFNISRYSQQLMETETSHR 1020
                *****

W3110_lacZ      HLLHAEEGTWLNIDGFHMETGIGGDSWSPSVSAEFQLSAGRYHYQLVWCQKSTOP 1076
lacZ-HT        HLLHAEEGTWLNIDGFHMETGIGGDSWSPSVSAEFQLSAGRYHYQLVWCQKSTOP 1076
                *****
```

