

# **Linking High Throughput Cell Culture, Multivariate Analysis and Economics for More Effective Process Integration**

A Thesis submitted to University College London for the degree of

Doctor of Philosophy (PhD) in Biochemical Engineering

By

**Alma Mona Antemie**

The Advanced Centre for Biochemical Engineering

Department of Biochemical Engineering

UCL

Gordon Street

London

WC1H 0AH

September 2015

I, *Alma Mona Antemie* confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

To my Mum,  
For always inspiring me to aim high

## Abstract

As the antibody sector has matured, it has seen significant increases in cell culture titres. However, it is hard to predict the consequences of titre increase on impurity levels and downstream processing (DSP) performance. Hence it is critical to have systematic methods to explore such interactions. This project explored the potential of high throughput cell culture linked to multivariate analysis, uncertainty analysis and bioprocess economics to characterise cell culture processes, not only in terms of growth and productivity but also host cell protein (HCP) levels, robustness and costs. A Quality by Design (QbD) approach to cell culture process development is presented. Using this QbD framework it was shown that there is scope for cell culture processes in which the ratio of mAb to HCP can be increased and the association of mAb titre to HCP reduced. It is therefore feasible to identify conditions whereby it is possible to increase antibody titre with little impact on HCP levels and hence subsequent DSP operations. (36.5 °C, 313 mOsm kg<sup>-1</sup> media osmolality, 1 × 10<sup>6</sup> cells mL<sup>-1</sup> seeding density, pH 6.8 and low cell generation number in this case). The impact of cell culture factors on protease activity (problematic HCP species) was assessed. Culture temperature was found to have a significant impact on protease activity, with a decrease in temperature resulting in lower protease activity. The relationship between HCP levels and protease activity was also examined and it was shown that an increase in total HCP levels at harvest did not result in a concomitant increase in protease activity. Multivariate data analysis based on regression was used to derive statistical cause-and-effect correlations able to link mAb titre and HCP levels to key cell culture factors. The resulting cell culture predictive correlations were then integrated into a whole bioprocess economics and optimisation framework. This allowed the identification of the most cost effective cell culture strategies as well as the impact of uncertainty in cell culture parameters on outputs (product output (kg) and HCP<sub>final</sub> (ng/mg)) and the likelihood of these falling out of specification. The work in this thesis highlights the benefits of a systematic approach to providing enhanced process understanding of the impact of cell culture strategies on downstream processes. This can be used to facilitate effective process integration and enable continuous improvements.

## **Acknowledgements**

I would like to thank my academic supervisors at UCL (Suzanne Farid and Gary Lye) and industrial supervisors at MedImmune (Steven Ruddock and Kenneth Lee) for their encouragement, patience and support during my PhD. The advice and help of Stephen Goldrick, Richard Allmendinger (UCL) as well as Ray Field, David Gruber, Jeremy Springall (MedImmune) have also been invaluable.

I would also like to thank MedImmune for providing the cell line used in this work and for their support with the Protein A and HCP Gyros assays. In addition I'd like to thank the upstream team at MedImmune for all their help and for always making me feel so welcomed during my time in Cambridge.

I am so grateful for all the love and support given by my family and friends. They have helped me overcome challenging and stressful times over the years.

Financial support from the Biotechnology and Biological Sciences Research Council (BBSRC) UK and MedImmune is gratefully acknowledged.

# Contents

<b>Abstract</b> .....	<b>4</b>
<b>Acknowledgements</b> .....	<b>5</b>
<b>Contents</b> .....	<b>6</b>
<b>List of Tables</b> .....	<b>12</b>
<b>List of Figures</b> .....	<b>15</b>
<b>Nomenclature</b> .....	<b>21</b>
<b>Chapter 1</b> .....	<b>25</b>
<b>1 Literature review</b> .....	<b>25</b>
<b>1.1 Biopharmaceutical drug development</b> .....	<b>26</b>
1.1.1 Marketed biopharmaceuticals .....	26
1.1.2 Marketed monoclonal antibodies .....	27
1.1.3 Development costs and risks .....	28
1.1.4 Manufacturing development and costs.....	29
<b>1.2 Antibody manufacture</b> .....	<b>34</b>
1.2.1 Antibody platform process .....	34
1.2.2 Quality by Design .....	37
1.2.3 Increases in mAb titres .....	38
1.2.4 Future trends of antibody platform.....	40
<b>1.3 Small scale cell culture</b> .....	<b>43</b>
1.3.1 Traditional small-scale culture systems .....	44

1.3.2	Shaken vessels.....	44
1.3.3	Stirred systems .....	47
<b>1.4</b>	<b>Impurities in biopharmaceutical manufacturing .....</b>	<b>50</b>
1.4.1	Host cell proteins (HCPs).....	51
1.4.2	Analytical tools for detection, monitoring and quantitation of HCPs .....	52
<b>1.5</b>	<b>Problematic HCPs .....</b>	<b>56</b>
1.5.1	Introduction .....	56
1.5.2	Common problematic HCPs.....	59
<b>1.6</b>	<b>Statistical modelling .....</b>	<b>61</b>
1.6.1	Introduction .....	61
1.6.2	Multivariate Data Analysis.....	62
1.6.3	Cell culture models .....	64
1.6.4	Predictive modelling .....	67
<b>1.7</b>	<b>Bioprocess economics modelling .....</b>	<b>68</b>
1.7.1	Introduction .....	68
1.7.2	Factors impacting COG/g.....	69
1.7.3	Classification of process economic models .....	70
1.7.4	Deterministic process economics .....	73
1.7.5	Stochastic process economics .....	75
1.7.6	Dynamic and stochastic studies.....	76
<b>1.8</b>	<b>Aims and Organisation of thesis.....</b>	<b>81</b>
<b>Chapter 2</b>	<b>.....</b>	<b>84</b>
<b>2</b>	<b>Materials and methods .....</b>	<b>84</b>

<b>2.1</b>	<b>Equipment used .....</b>	<b>84</b>
2.1.1	ambr system .....	84
<b>2.2</b>	<b>Cell culture .....</b>	<b>84</b>
2.2.1	Cell line .....	84
2.2.2	Fed-batch protocols.....	85
<b>2.3</b>	<b>DoE experiments.....</b>	<b>86</b>
2.3.1	AMBR 1 .....	86
2.3.2	AMBR 2 .....	86
2.3.3	AMBR 3 .....	87
2.3.4	AMBR 4.....	87
<b>2.4</b>	<b>Analytical techniques .....</b>	<b>87</b>
2.4.1	CCH, $q_{\text{Mab}}$ and specific HCP productivity ( $q_{\text{HCP}}$ ).....	87
2.4.2	Cell generation number .....	88
2.4.3	Viable cell concentration and Viability.....	89
2.4.4	Protein A HPLC Analysis .....	89
2.4.5	Gyroslab™ xP workstation .....	89
2.4.6	Protease assay.....	90
<b>2.5</b>	<b>Statistical analysis.....</b>	<b>92</b>
2.5.1	Spearman’s Rank correlation test.....	92
2.5.2	Model building .....	92
2.5.3	Model selection .....	94
2.5.4	Model assumptions.....	95
2.5.5	Model validation .....	96

2.5.6	Definitions.....	97
<b>2.6</b>	<b>Software used .....</b>	<b>99</b>
2.6.1	JMP and Design Expert.....	99
2.6.2	C#, Microsoft Visual Studio and Microsoft Access.....	100
<b>Chapter 3</b>	<b>.....</b>	<b>101</b>
<b>3</b>	<b>Quality by design approach to cell culture process development.....</b>	<b>101</b>
<b>3.1</b>	<b>Introduction .....</b>	<b>101</b>
<b>3.2</b>	<b>Reproducibility of ambr system .....</b>	<b>101</b>
<b>3.3</b>	<b>Growth profiles of CHO cells in fed-batch culture.....</b>	<b>103</b>
<b>3.4</b>	<b>Effect of cell culture inputs and generation number on cell growth.....</b>	<b>104</b>
<b>3.5</b>	<b>Effect of cell culture inputs and generation number on <math>q_{Mab}</math>, titre and HCP levels.....</b>	<b>111</b>
3.5.1	$q_{Mab}$ .....	111
3.5.2	mAb titre and HCP levels.....	115
<b>3.6</b>	<b>Analysis.....</b>	<b>119</b>
3.6.1	Analysis of low generation cell experiments (AMBR 1) .....	120
3.6.2	Analysis of mid generation cell experiments (AMBR 2) .....	121
3.6.3	Analysis of high generation cell experiments (AMBR 3) .....	121
<b>3.7</b>	<b>Conclusions .....</b>	<b>126</b>
<b>Chapter 4</b>	<b>.....</b>	<b>127</b>
<b>4</b>	<b>Analysis of problematic HCPs in cell culture processes .....</b>	<b>127</b>
<b>4.1</b>	<b>Introduction .....</b>	<b>127</b>
<b>4.2</b>	<b>Protease assay development.....</b>	<b>127</b>
4.2.1	Trypsin standard curve .....	127

4.2.2	Investigate the colour effect of unknown samples .....	128
4.2.3	Effect of unclarified material on assay performance.....	131
<b>4.3</b>	<b>Analysis of unknown samples .....</b>	<b>133</b>
4.3.1	Identification of suitable dilutions.....	133
4.3.2	Impact of cell culture inputs on protease activity.....	136
4.3.3	Correlation between mAb titre, HCP levels and protease activity .....	142
<b>4.4</b>	<b>Conclusions .....</b>	<b>146</b>
<b>Chapter 5</b>	<b>.....</b>	<b>148</b>
<b>5</b>	<b>Derivation of predictive cause-and-effect correlations for cell culture using multivariate data analysis .....</b>	<b>148</b>
<b>5.1</b>	<b>Introduction .....</b>	<b>148</b>
<b>5.2</b>	<b>Results and discussions .....</b>	<b>148</b>
5.2.1	Titre prediction models from AMBR runs .....	149
5.2.2	HCP prediction models from AMBR runs .....	162
5.2.3	Importance of predictive models within QbD .....	169
<b>5.3</b>	<b>Conclusions .....</b>	<b>169</b>
<b>Chapter 6</b>	<b>.....</b>	<b>171</b>
<b>6</b>	<b>Integration of predictive cell culture correlations with bioprocess economics and uncertainty analysis .....</b>	<b>171</b>
<b>6.1</b>	<b>Introduction .....</b>	<b>171</b>
<b>6.2</b>	<b>Methodology.....</b>	<b>173</b>
6.2.1	Problem definition.....	173
<b>6.3</b>	<b>Framework description.....</b>	<b>174</b>
6.3.1	Bioprocess economics model.....	175

<b>6.4</b>	<b>Case study description.....</b>	<b>176</b>
<b>6.5</b>	<b>Results and discussions .....</b>	<b>176</b>
6.5.1	Impact of cell culture parameters on COG/g.....	176
<b>6.6</b>	<b>Conclusions .....</b>	<b>194</b>
<b>Chapter 7</b>	<b>.....</b>	<b>196</b>
<b>7</b>	<b>Conclusions and Future work.....</b>	<b>196</b>
7.1	Conclusions .....	196
7.2	Future work .....	198
<b>References</b>	<b>.....</b>	<b>202</b>
<b>Appendix</b>	<b>.....</b>	<b>225</b>
<b>Chapter 5 Appendix</b>	<b>.....</b>	<b>225</b>

## List of Tables

Table 1-1: Top 10 best-selling biotech drugs in 2014*.....	31
Table 1-2: FDA approved antibody-based therapeutics*.....	32
Table 1-3: Capital investment costs for antibody-based facilities using mammalian cells* .....	33
Table 1-4: Analysis of different production processes for marketed antibody-based therapeutics* .....	37
Table 1-5: Classification system of mathematical kinetic cell culture models* .....	65
Table 3-1: Experimental design setup for AMBR 1, AMBR 2 and AMBR 3 experiments .....	107
Table 3-2: The effect of generation number on CCH viable cells, $q_{\text{Mab}}$ , antibody titre and HCP levels. The comparison is done between replicate conditions of AMBR 2 and AMBR 3 in comparison to AMBR 1.....	113
Table 3-3: Summary of Spearman’s rank test for the correlations between variables in Figure 3-5.....	124
Table 3-4: Summary of Spearman’s rank test for the correlations between variables in Figure 3-7.....	125
Table 5-1: Summary statistics for eight “best” models for AMBR 1 titre presented in coded factors. The coded factors A, B and C refer to the cell culture variables media osmolality, seeding density and temperature, respectively .....	153
Table 5-2: Comparison between generated titre models and titre models found in literature. The comparison looks at the main parameters having a significant impact on product titre as well as $R^2$ and $R^2$ predicted.....	156
Table 5-3: The statistics for the Shapiro – Wilk and Durbin – Watson Tests and their corresponding $p$ -value for AMBR 1 Titre and HCP .....	158

Table 5-4: Summary statistics for eight “best” models for AMBR 1 HCP presented in coded factors. The coded factors A, B and C refer to the cell culture variables media osmolality, seeding density and temperature, respectively .....	160
Table 5-5: Comparison between generated HCP models and HCP models found in literature. The comparison looks at the main parameters having a significant impact on product titre as well as $R^2$ and $R^2$ predicted .....	165
Table 5-6: Final predictive models for titre and HCP from each AMBR run presented in coded factors. The coded factors A, B, C, D and E refer to the cell culture variables media osmolality, seeding density, temperature, pH and feed start time, respectively .....	168
Table 6-1: The experimental design investigated for the input cell culture parameters within AMBR 1, 2 and 3 .....	177
Table 6-2: Details of case study scenario (Panel A) and probability distributions of uncertain factors (Panel B)* .....	178
Table 6-3: Combination of input cell culture parameters for AMBR 1, 2 and 3 with their associated Prot A and CEX column sizing, for the lowest and highest COG/g achieved in each case.....	182
Table 5-1A: Summary statistics for eight “best” models for AMBR 2 titre presented in coded factors. The coded factors B, D and E refer to the cell culture variables seeding density, pH and feed start time, respectively .....	226
Table 5-2A: Summary statistics for eight “best” models for AMBR 2 HCP presented in coded factors. The coded factors B, D and E refer to the cell culture variables seeding density, pH and feed start time, respectively. ....	227
Table 5-3A: Summary statistics for eight “best” models for AMBR 3 titre presented in coded factors. The coded factors A, B and C refer to the cell culture variables media osmolality, seeding density and temperature, respectively. ....	228

Table 5-4A: Summary statistics for eight “best” models for AMBR 3 HCP presented in coded factors. The coded factors A, B and C refer to the cell culture variables media osmolality, seeding density and temperature, respectively. .... 229

## List of Figures

Figure 1-1 : Typical upstream process for monoclonal antibodies .....	35
Figure 1-2: Typical downstream process for monoclonal antibodies .....	35
Figure 1-3: Micro 24 bioreactor .....	46
Figure 1-4: Illustration of (a) ambr microscale bioreactor vessel (b) automated ambr workstation.....	48
Figure 1-5: Overview of deterministic and stochastic bioprocess economic models presented in this section .....	72
Figure 3-1: Fed-batch culture profiles for centre points within AMBR 1 A) viable cell density, B) glucose concentration, C) lactate concentration. The cultures were seeded at $0.8 \times 10^6$ cells ml <sup>-1</sup> viable cell density, 353 mOsm kg <sup>-1</sup> and 36.5 °C.....	102
Figure 3-2: Fed-batch culture profiles for AMBR 1 A) viable cell density, B) viability profiles, C) antibody titre. Two different temperatures have been used - - 36.5 °C, ... 33 °C. Within the DoE design, cultures were seeded at a density of 0.49, 0.8, $1.14 \times 10^6$ cells mL <sup>-1</sup> and the osmolality levels used were 314, 353, 394 mOsm kg <sup>-1</sup> respectively. The pH was maintained at $6.8 \pm 0.1$ and feeding started on day 2 of culture.....	108
Figure 3-3: Fed-batch culture profiles for AMBR 2 A) viable cell density, B) viability profiles, C) antibody titre. Three different pH levels have been used ... 6.6, - 6.8 and 7. Within CS1 of AMBR 2 cultures, seeding density and pH levels were 0.42, 0.69, $1.06 \times 10^6$ cells mL <sup>-1</sup> and 6.6, 6.8, 7 while feeding started on day 2 of culture. Within CS2 of AMBR 2 cultures, pH and timing of feed initiation levels were 6.6, 6.8, 7 and day 1, 2, 3 respectively, while the starting seeding density was kept constant at $1 \times 10^6$ cells mL <sup>-1</sup> . A temperature of 36.5 °C and an osmolality of 313 mOsm kg <sup>-1</sup> were maintained for all 24 cultures.....	109

Figure 3-4: Fed-batch culture profiles for AMBR 3 A) viable cell density, B) viability profiles, C) antibody titre. Two different temperatures have been used - - 36.5 °C, ··· 33 °C. Within the DoE design, cultures were seeded at a density of 0.58, 1.1, 1.8 × 10<sup>6</sup> cells mL<sup>-1</sup> and the osmolality levels used were 320, 350, 390 mOsm kg<sup>-1</sup> respectively. The pH was maintained at 6.8 ± 0.1 and feeding started on day 2 of culture..... 110

Figure 3-5: The plot of A), B), C) viability against HCP levels and the observed correlations between D), E), F) q<sub>Mab</sub> and q<sub>HCP</sub> within AMBR 1, AMBR 2 and AMBR3 experiments, respectively. See Materials and methods for details on AMBR 1, AMBR 2 and AMBR 3 experimental designs..... 114

Figure 3-6: Impact of different cell culture parameters (seeding density, media osmolality and timing of feed initiation) on antibody titre and HCP levels at two different temperatures ··· 36.5 °C, ··· 33 °C, within AMBR 1, AMBR 2 and AMBR 3 experiments. See Materials and methods for details on AMBR 1, AMBR 2 and AMBR 3 experimental designs. .... 116

Figure 3-7: A correlation plot of antibody titre against HCP levels within AMBR 1, AMBR 2 and AMBR 3 experiments. The relationship between A) antibody titre, CCH and q<sub>Mab</sub>, B) antibody titre and HCP levels, C) CCH viable cells and q<sub>Mab</sub>: q<sub>HCP</sub> ratio is shown. See Materials and methods for details on AMBR 1, AMBR 2 and AMBR 3 experimental designs. .... 123

Figure 4-1: Trypsin standard curve. The trypsin concentration varied from 50 – 10,000 ng mL<sup>-1</sup>, in-house Tris-HCl digestion buffer was used for making up the dilutions and the plate was incubated in the dark at 40° C and measured at hourly intervals. The increase in trypsin concentration is proportional to the increase in fluorescence ..... 128

Figure 4-2: Investigation into the colour effect of unknown samples. A) Fluorescence increase associated with the unknown sample containing the substrate (U+S) (—) and substrate-free control (U+B) (···), B) Fluorescence increase due to protease activity (U+S) – (U+B)..... 131

- Figure 4-3: Comparison between unclarified and clarified levels of the same unknown sample. Clarification was achieved through either centrifugation ( — ) or centrifugation followed by sterile filtration ( - - ). A) Fluorescence increase associated with the three levels of clarification of the unknown sample containing the substrate (U+S) ( — ) and substrate-free control (U+B) ( - - - ), B) Fluorescence increase due to protease activity (U+S) – (U+B), for each level of clarification ..... 133
- Figure 4-4: Protease activity increase over time for an unknown sample. A) The fluorescence increase of the sample with substrate (U+S) ( — ) and substrate-free sample (U+B) ( - - - ) for dilutions 1/10 ( — ), 1/50 ( - - ) and 1/100 ( - - - ). B) Fluorescence increase due to protease activity ((U+S) – (U+B)) for each of the three dilutions..... 135
- Figure 4-5: Trypsin standard curve. The trypsin concentration varied from 50 – 10,000 ng mL<sup>-1</sup>, in-house Tris-HCl digestion buffer was used for making up the dilutions and the plate was incubated in the dark at 40° C. The increase in trypsin concentration is proportional to the increase in fluorescence ..... 136
- Figure 4-6: Difference in protease activity (%) between equivalent conditions at 33 °C compared to 36.5 °C for (A) AMBR 3 and (B) AMBR 4 ..... 138
- Figure 4-7: The impact of seeding density on protease activity (measured as a fluorescence increase) within AMBR 3 (A) and AMBR 4 (B) cultures ..... 140
- Figure 4-8: The impact of starting media osmolality on protease activity (measured as a fluorescence increase) within AMBR 3 (A) and AMBR 4 (B) cultures... 141
- Figure 4-9: The relationship between protease activity (measured as a fluorescence increase) and mAb titre within AMBR 3 (A) and AMBR 4 (B) cultures ..... 143
- Figure 4-10: The relationship between protease activity (measured as a fluorescence increase) and mAb titre within AMBR 4, split by pH. A) 6.8 B) 7-7.4..... 144

Figure 4-11: The relationship between protease activity (measured as a fluorescence increase) and HCP levels within AMBR 3 (A) and AMBR 4 (B) cultures .....	146
Figure 5-1: Integration of AMBR run experiments with predictive modelling .....	151
Figure 5-2: Analysis of mAb titre models for AMBR 1, 2 and 3. The analysis of variance as well as the correlation between experimental values and predictions from each model is presented * Significant in terms of probability (p-value < 0.05) .....	152
Figure 5-3: Comparison between observed and predicted values of AMBR 1 titre. Runs 1-11 correspond to CS1 within AMBR 1, maintained at a temperature of 36.5 °C and runs 11-20 correspond to CS2 within AMBR 1, maintained at a temperature of 36.5 °C followed by a temperature shift to 33 °C in day 4 of culture.....	157
Figure 5-4: Residual versus predicted values AMBR 1 titre .....	159
Figure 5-5: Analysis of HCP models for AMBR 1, 2 and 3. The analysis of variance as well as the correlation between experimental values and predictions from each model is presented .....	161
Figure 5-6: Comparison between observed and predicted values of AMBR 1 HCP. Runs 1-11 correspond to CS1 within AMBR 1, maintained at a temperature of 36.5 °C and runs 11-20 correspond to CS2 within AMBR 1, maintained at a temperature of 36.5 °C followed by a temperature shift to 33 °C in day 4 of culture.....	166
Figure 5-7: Residual vs predicted values AMBR 1 HCP.....	167
Figure 6-1: Overall integration of HT experimentation, predictive (statistical) modelling and economic modelling .....	172
Figure 6-2: Main components of integrated bioprocess economics and optimisation tool* .....	175

Figure 6-3: The impact of titre and HCP load resulting from AMBR 1 cell culture at A) 36.5 °C and B) 33 °C on the COG/g of a fixed USP and DSP process.....	180
Figure 6-4: COG breakdown for the lowest and highest COG/g achieved within the design space of AMBR 1 .....	181
Figure 6-5: The impact of cell culture inputs on mAb titre and HCP levels within AMBR 2 cultures .....	183
Figure 6-6: The impact of cell culture inputs on COG/g within AMBR 2 cultures	184
Figure 6-7: The impact of titre and HCP load resulting from AMBR 3 cell culture at A) 36.5 °C and B) 33 °C on the COG/g of a fixed USP and DSP process.....	185
Figure 6-8: Boxplots showing the distribution of COG/g for different AMBR runs. The box represents the 25 <sup>th</sup> and 75 <sup>th</sup> percentile with the median indicated by the middle horizontal line. The whiskers represent the observations with the lowest and highest COG/g for each AMBR run.....	186
Figure 6-9: Running average in product yield for A) AMBR 1, C) AMBR 3 and in HCP final for B) AMBR 2 during Monte Carlo simulation trials. ....	188
Figure 6-10: Steps used for establishing PARs/PVAC using HT experimentation, ANOVA and Monte Carlo simulation methods.....	191
Figure 6-11: Density plots showing the impact of tighter osmolality and seeding density control on the probability (P(x)) of the product output falling out of specification in AMBR 1 and AMBR 3, using MC = 150 for Seeding density $0.75 \pm 0.25 \times 10^6$ cells/mL A) Osmolality $350 \pm 40$ mOsm/kg; B) Osmolality $340 \pm 30$ mOsm/kg; C) Osmolality $330 \pm 20$ mOsm/kg in AMBR 1 and for Seeding density $1.2 \pm 0.6 \times 10^6$ cells/mL D) Osmolality $350 \pm 40$ mOsm/kg; E) Osmolality $340 \pm 30$ mOsm/kg; F) Seeding density $1.2 \pm 0.2 \times 10^6$ cells/mL, Osmolality $340 \pm 30$ mOsm/kg in AMBR 3 .....	192

Figure 6-12: Density plots showing the impact of tighter pH and seeding density control on the probability (P(x)) of the HCP<sub>final</sub> (ng/mg) falling out of specification in AMBR 2, using MC = 150 for Seeding density  $0.75 \pm 0.25 \times 10^6$  cells/mL A) pH  $6.8 \pm 0.2$ ; B) pH  $6.8 \pm 0.1$ ; C) Seeding density  $0.75 \pm 0.2 \times 10^6$  cells/mL, pH  $6.8 \pm 0.1$ ..... 193

## **Nomenclature**

<b>AEX</b>	Anion exchange
<b>AIC</b>	Akaike information criterion
<b>ambr</b>	Advanced micro-scale bioreactor
<b>ANOVA</b>	Analysis of variance
<b>ATF</b>	Alternative tangential flow
<b>BIC</b>	Bayesian information criterion
<b>CAGR</b>	Compound annual growth rate
<b>CCH</b>	Cumulative cell hours
<b>CEX</b>	Cation exchange
<b>CFD</b>	Computational fluid dynamics
<b>CHO</b>	Chinese hamster ovary
<b>COG</b>	Cost of goods
<b>CQA</b>	Critical quality attribute
<b>CS</b>	Culture station
<b>CV</b>	Cross validation
<b>DO</b>	Dissolved oxygen
<b>DoE</b>	Design of experiments
<b>DNA</b>	Deoxyribonucleic acid
<b>DSP</b>	Downstream processing
<b>EA</b>	Evolutionary algorithm
<b>ECS</b>	Enhanced cell settling
<b>ELISA</b>	Enzyme-linked immunosorbent assay
<b>EMOA</b>	Evolutionary multiobjective optimization algorithm
<b>FCI</b>	Fixed capital investment

<b>FDA</b>	Food and drug administration
<b>FT-MIR</b>	Fourier transform mid infrared spectrometry
<b>GMP</b>	Good manufacturing practice
<b>HAMA</b>	Anti-murine antibodies
<b>HCCF</b>	Harvested cell culture fluid
<b>HCP</b>	Host cell protein
<b>HPLC</b>	High pressure liquid chromatography
<b>HT</b>	High throughput
<b>HTS</b>	High throughput screening
<b>HVAC</b>	Heating, ventilation and air-conditioning
<b>IEX</b>	Ion exchange
<b>LRV</b>	Log10 reduction value
<b>M24</b>	Micro 24
<b>mAb</b>	Monoclonal antibody
<b>MB</b>	Micro bioreactor
<b>MBA</b>	Micro bioreactor array
<b>MC</b>	Monte Carlo
<b>MLR</b>	Multiple linear regression
<b>MTP</b>	Microtitre plate
<b>MVDA</b>	Multivariate data analysis
<b>NPV</b>	Net present value
<b>OD</b>	Optical density
<b>OTR</b>	Oxygen mass transfer rate
<b>pAbs</b>	Polyclonal antibodies
<b>PAR</b>	Proven acceptable range
<b>PCA</b>	Principal component analysis

<b>PLS</b>	Partial least squares
<b>PRESS</b>	Predicted residual sum of squares
<b>PVAC</b>	Process validation acceptance criteria
<b>QbD</b>	Quality by Design
<b>q<sub>HCP</sub></b>	Specific host cell protein productivity
<b>q<sub>Mab</sub></b>	Specific monoclonal antibody productivity
<b>q<sub>p</sub></b>	Specific cell productivity
<b>r</b>	Pearson's coefficient
<b>r<sub>s</sub></b>	Spearman's rank
<b>R</b>	Regression coefficient
<b>R<sup>2</sup></b>	Coefficient of determination
<b>R&amp;D</b>	Research and development
<b>RMSE</b>	Root mean squared error
<b>SP<sub>XY</sub></b>	Sum of products
<b>SSE</b>	Squared sum of prediction errors
<b>SS<sub>X</sub></b>	X sum of squares
<b>SS<sub>Y</sub></b>	Y sum of squares
<b>TFF</b>	Tangential flow filtration
<b>USP</b>	Upstream processing
<b>VCD</b>	Viable cell density
<b>VCN</b>	Viable cell number
$\bar{X}$	Mean of X
<b>XD</b>	Extreme density
<b>X<sub>model</sub></b>	Predicted values of X
<b>X<sub>obs</sub></b>	Observed values of X
$\bar{Y}$	Mean of Y

$Y_{\text{model}}$	Predicted values of Y
$Y_{\text{obs}}$	Observed values of Y
w/v	Working volume

# Chapter 1

## 1 Literature review

As the antibody sector has matured, it has seen significant increases in upstream (USP) productivities that have opened up the possibility for radical changes to the design and operation of cell culture suites. However, due to the inherently complex set of interactions that can affect cell culture performance, it is hard to predict the consequences on the impurity profiles and hence robustness of downstream (DSP) operations as titre increases. Quality by Design (QbD) initiatives are driving the need for greater understanding of the impact of USP changes on DSP, such as the impact of cell culture strategies on the downstream processing equipment duties so as to enable effective process integration and hence continuous improvements.

This project will explore state-of-the-art high throughput cell culture and multivariate data analysis techniques to characterise cell culture operations, not only in terms of growth and productivity but also impurity levels. The resulting cell culture statistical cause-and-effect correlations will be integrated into process economics models so as to identify the most cost-effective integrated USP and DSP manufacturing strategies for the future.

This introductory chapter provides an overview of biopharmaceutical drug development, impurities and the challenges they pose within the manufacturing process as well as outlining statistical and bioprocess economics modelling techniques. Section 1.1 provides an overview of biopharmaceutical drug development and its associated costs and risks. Section 1.2 reviews the current and future trends of monoclonal antibody manufacturing. Section 1.3 gives an outline of existing small-scale cell culture systems. Sections 1.4 and 1.5 focus on impurities present in biopharmaceutical manufacturing and the challenges they pose, with an emphasis on host cell proteins. Sections 1.6 and 1.7 highlight statistical and economics modelling techniques employed in evaluating biopharmaceutical manufacturing processes. Finally, the aim and organisation of the thesis are presented in Section 1.8.

## **1.1 Biopharmaceutical drug development**

Biopharmaceuticals are being developed to target different types of illnesses such as diabetes, multiple sclerosis, cancer, hepatitis and viral infections. They can be split into several categories: enzymes, hormones, antisense drugs, vaccines, monoclonal antibodies, cytokines, clotting factors, peptide therapeutics and cell therapies (Sekhon, 2010). In this thesis the focus will be towards monoclonal antibodies (mAbs) as they have become essential to current medicine and represent one of the main biopharmaceutical products in development at the moment (Rodrigues et al., 2010; Ecker et al., 2015).

### **1.1.1 Marketed biopharmaceuticals**

Biologics are increasingly important and now represent some 50 % of the portfolio of major pharmaceutical companies. Every year there are more than 600 biologics under development, the majority of which are monoclonal antibodies (Van Amum, 2015). The first recombinant protein approved on the market was genetically engineered insulin in 1982. Following insulin's great success on the market, the US Food and Drug Administration (FDA) has approved more than 300 non-recombinant therapeutics, including vaccines and blood products as well as over 100 recombinant protein biopharmaceuticals ([www.fda.gov](http://www.fda.gov); Rader, 2013).

Mammalian cell expression has become the predominant production system for recombinant proteins due to its ability to carry out human-like post-translational modifications and to synthesize proteins very similar to those occurring naturally in terms of biochemical properties and molecular structure (Zhu, 2011, Khan, 2013). Almost all of the biopharmaceuticals approved are made in either mammalian cells, bacterium (*E.coli*) or yeast (*Pichia pastoris*, *Saccharomyces cerevisiae*) (Berlec and Strukelj, 2013). Around 70 % of approved recombinant proteins are expressed in CHO cells (Gutierrez et al., 2012). This shows that mammalian expression systems are the main choice for biopharmaceutical manufacturing.

With regards to the economic impact of marketed top selling biopharmaceutical products (Table 1-1), almost 75 % of the total revenue comes

from products manufactured using mammalian cell expression systems and the remaining 25 % from products manufactured using bacterial (e.g. *E.coli*). The worldwide market of therapeutic proteins was valued at \$93 billion in 2010 and is predicted to increase to \$141.5 billion by 2017. This constitutes an increase rate of 6.2 % within the 7 years period (GBI Research, 2011).

### **1.1.2 Marketed monoclonal antibodies**

Monoclonal antibodies (mAbs) represent a considerable proportion of the biopharmaceutical industry's current biotherapeutics portfolio (Zhang et al., 2014). As a result of their extensive application range, mAbs are globally used in a wide range of applications such as therapeutics or diagnostics. In therapeutics they are used to treat a wide range of disease conditions, with autoimmune disorders and different types of cancer being the most frequent targets (Table 1-2) (Shukla and Thommes, 2010; Elvin et al., 2013).

The majority of mAbs on the market and in clinical development are derived from a few mammalian cell lines: Sp2/0, NS0, CHO (Chinese Hamster Ovary). CHO cells are the most commonly used expression system for the industrial production of recombinant proteins, including monoclonal antibodies. Some of the key advantages of CHO include the fact that methods for gene amplification, cell transfection and clone selection are well characterized and the protein of interest is secreted in the culture media which simplifies the purification process (Gutierrez et al., 2012; Bailey-Kellogg et al., 2014).

The mAbs market experienced newfound interest at the turn of the century and has been steadily increasing over the past decade. Antibodies are the fastest growing segment of biopharmaceuticals; in 2006, the global mAbs market was estimated to be \$20.6 billion with only 25 approved mAbs on the market (Coco-Martin and Hamsen, 2008), in 2008 to be \$39 billion while in 2013 this grew to \$75 billion, and 47 mAb products approved in the US and Europe, almost double compared to 2006. The mAbs market has seen a 90 % increase in growth sales between 2008 and 2013, compared to 26 % for sale of other recombinant protein

therapeutics. Out of the 47 mAb products approved in 2013, 18 of them had sales over \$1 billion, 7 had sales over \$6 billion and Humira (AbbVie Inc.) reported record sales of ~ \$12.5 billion in 2014. Their success will continue to dominate the biopharmaceutical pipeline and market, with analysts forecasting a compound annual growth rate of 3-5 new products approved/year, leading to 70 mAb products on the market by 2020 and combined world sales of nearly \$125 billion (Ecker et al., 2015). The main drivers behind the continuing increase of mAb products on the market are the ageing worldwide population as well as an increase in standard of living in emerging markets.

### **1.1.3 Development costs and risks**

The development of therapeutic antibodies is a long, expensive and risky process that has to comply with rigorous regulatory requirements. Tough decisions must be made in order to find a balance between speed to market, low cost, high quality and flexibility (Farid, 2009). Biopharmaceutical companies therefore need to achieve development efficiency and an advantage over competitors by starting to use new bioprocess technologies (Bareither and Pollard, 2011). Drug supply development cost needs to be lowered and the time from discovery to market must also be reduced. To develop and bring a new drug to the market takes up to 10-15 years (Roy, 2011; Schnatz, 2013). The estimate for drug development costs was \$802m in 2003 (DiMasi et.al, 2003), over \$1 billion in 2010 (Adams and Brantner, 2010) and \$2.6 billion in 2015 (Avorn, 2015). Due to FDA's strict regulatory requirements clinical trials have low success rates (~16 % from Phase I to market). On average, seven years are needed to go through the clinical trials and obtain the regulatory approvals, to market a new drug (Kaitin, 2010).

Clinical trials are made up of three phases. In Phase I, the required drug is tested for safety on a small number of healthy volunteers (20-80), in Phase II, the drug is tested in terms of safety and efficacy on a small number of patients (100-300) and Phase III involves much larger scale clinical studies that monitors the long term use of the drug with up to 3000 patients participating. They account for most of the drug development cost, with reports showing up to 70 % of all research and

development (R&D) costs being spent on Phase I to Phase IV clinical trials (Bernt and Cockburn, 2012). A high number of candidates must be screened in order to pick out a small number of prospective drugs. From these, only a few are found to be effective at treating diseases. For a company to remain successful, it needs to have a selection of drugs in the pipeline. Using an economic model of drug discovery and development, Paul et al. (2010) predicted that large Pharma companies should aim for 2-5 launches per year, requiring 18-45 molecules to enter Phase I clinical trials. These numbers are rarely achieved, even by very large companies. A more realistic estimation would be 1-2 launches per year, requiring around 9-18 molecules to enter Phase I clinical trials (Farid, 2001). The success rate of bringing a new drug to the market has lowered since the FDA has become stricter regarding the approval processes. In 2007, FDA approved only 19 new drugs, representing the lowest number since 1983 (Courtenay, 2008).

#### **1.1.4 Manufacturing development and costs**

Monoclonal antibodies have relatively low potency and require high doses over an extended period of time, which involves large amounts of purified product per patient (Aldington and Bonnerjea, 2007; Tao et al., 2014). Antibody manufacturing facilities now reach sizes of around 45,000 m<sup>2</sup> and have multiple bioreactors with total capacities up to 200,000 L (Farid, 2009).

One of the major challenges of biopharmaceuticals, especially monoclonal antibodies is reducing the selling price in order to make them more affordable to patients. Monoclonal antibodies based therapeutics for autoimmune diseases cost around \$15,000 - \$20,000 per year to treat a single patient. Products designed for rare diseases will cost a lot more. The reason why these therapies are so expensive is partly due to the need for frequent administration of high doses. It is also due to the high drug development costs given the high attrition rates that mean companies need to recoup the investment made in the development of not only the successful drug but also the failed drugs. A major component of the drug development process is the establishment of robust manufacturing processes and analytical methods for each drug candidate in the portfolio. Monoclonal antibodies are manufactured in

mammalian cells, involving a time-consuming two weeks of cell culture to obtain the product. Monoclonal antibody based drugs also require frequent injections at high doses to be effective. Taking Humira, a mAb based drug to treat rheumatoid arthritis, as an example, the recommended dose is 40 milligrams every two weeks, with a total of over 1 g per year, per patient (Fanneau de la Horie, 2010).

Fixed capital investment (FCI) represents the amount of money required to build the biopharmaceutical facility ready to use. It generally involves the construction expenses for the buildings, the cost of purchased equipment, piping and instrumentation and utilities. Investment costs for monoclonal antibodies manufacturing facilities range between \$40 M to \$650 M with \$7,130 - \$17,000 per m<sup>2</sup> and \$1,765-\$4,220 per L for volumes between 20,000-200,000 L (Farid, 2007). Table 1-3 shows the capital investment costs reported for facilities built by large biopharmaceutical companies such as Genentech and Lonza. Table 1-3 also shows that these facilities reach sizes of 500,000 ft<sup>2</sup> (46,450 m<sup>2</sup>) with bioreactor volumes of up to 200,000L attained with multiple bioreactors (Farid, 2007).

The biopharmaceutical industry is facing some difficult challenges such as pressure from healthcare providers to decrease drug prices, reduced average patent life, increased competition in generic markets and a reduction in revenues due to patent expiration (Collier, 2009). Biopharmaceutical manufacturing involves advanced technology and very rigorous regulatory compliance such as good manufacturing practices (GMP). Pharmaceutical companies are under high pressure to minimise fixed costs so they resort to reducing their internal capacities in manufacturing and R&D and increase their outsourcing (Zhang et al., 2011). The current total global pharmaceutical outsourcing market is around \$130 billion. From 2009-2014 it grew at a compound annual growth rate (CAGR) of 9.4 % and is further expected to grow at a CAGR of 8.7 % from 2015-2020. It is predicted to reach up to \$215 billion by 2020 (PR Newswire Research, 2015).

**Table 1-1: Top 10 best-selling biotech drugs in 2014\***

<b>Product</b>	<b>Revenue US \$ Billion (B)</b>	<b>Date approved</b>	<b>Manufacturer</b>	<b>Expression system</b>	<b>Indication</b>
Humira	12.54	2002	AbbVie	Mammalian (CHO)	Rheumatoid arthritis,
Remicade	9.25	1998	Johnson & Johnson, Merck Co.	Mammalian (NSO)	Crohn disease
Rituxan	8.68	1997	Roche, Biogen Idec	Mammalian (CHO)	Non-Hodgkin lymphoma
Enbrel	8.54	1998	Amgen, Pfizer	Mammalian (CHO)	Rheumatoid arthritis
Lantus	7.28	2000	Sanofi	<i>E.coli</i>	Diabetes
Avastin	6.96	2004	Roche	Mammalian (CHO)	Colorectal cancer
Herceptin	6.79	1998	Roche	Mammalian (CHO)	Breast cancer
Neulasta	5.86	1998	Amgen, Kyowa Hakko Kirin	<i>E.coli</i>	Myelosuppressive Chemotherapy
Prevnar	4.46	2002	Pfizer	Bacterial	Prevention of invasive pneumococcal disease
Avonex	3.01	1996	Biogen Idec	Mammalian (CHO)	Multiple sclerosis

\* Adapted from <http://cellculturedish.com/2015/03/10-biologics-on-best-selling-drugs-list-for-2014/>

**Table 1-2: FDA approved antibody-based therapeutics\***

<b>Brand Name: INN</b>	<b>Target: Antibody type</b>	<b>Indication</b>	<b>Company</b>	<b>Approval Date</b>
(Pending): Idarucizumab	Dabigatran: Humanized Fab	Anticoagulation	Boehringer Ingelheim	Pending
(Pending): Mepolizumab	IL-5: Humanized IgG1	Asthma	GlaxoSmithKline	Pending
(Pending): Necitumumab	EGFR: Human IgG1	Cancer	Eli Lilly & Co.	Pending
Repatha: Evolocumab	PCSK9: Human IgG2	High cholesterol	Amgen	Pending
Praluent: Alirocumab	PCSK9: Human IgG1	High cholesterol	Sanofi and Regeneron Pharmaceuticals	2015 (US)
Unituxin: Dinutuximab	GD2: Chimeric IgG1	Cancer	United Therapeutics	2015 (US)
Cosentyx: Secukinumab	IL-17a: Human IgG1	Autoimmune	Novartis	2015 (US,EU)
Opdivo: Nivolumab	PD1: Human IgG4	Cancer	Bristol-Myers Squibb	2014 (US)
Blinicyto: Blinatumomab	CD19, CD3: Murine bispecific tandem scFv	Cancer	Amgen	2014 (US)
Keytruda: Pembrolizumab	PD1: Humanized IgG4	Cancer	Merck & Co.	2014 (US)
Cyramza: Ramucirumab	VEGFR2: Human IgG1	Cancer	Eli Lilly & Co.	2014 (US,EU)
Entyvio: Vedolizumab	$\alpha 4\beta 7$ integrin: Humanized IgG1	Autoimmune	Takeda Pharmaceuticals U.S.A.	2014 (US,EU)
Sylvant: Siltuximab	IL-6: Chimeric IgG1	Castleman disease	Janssen Biotech	2014 (US,EU)
Gazyva: Obinutuzumab	CD20: Humanized IgG1, glycoengineered	Cancer	Genentech	2013 (US) 2014 (EU)
Kadcyla: Ado-trastuzumab emtansine	HER2: Humanized IgG1, immunoconjugate	Cancer	Genentech	2013 (US,EU)
Abthrax: Raxibacumab	B. anthraxis PA: Human IgG1	Anti-infective	Human Genome Sciences	2012 (US)
Perjeta: Pertuzumab	HER2: Humanized IgG1	Cancer	Genentech	2012 (US) 2013 (EU)
Adcetris: Brentuximab vedotin	CD30: Chimeric IgG1, immunoconjugate	Cancer	Seattle Genetics	2011 (US) 2012 (EU)
Benlysta: Belimumab	BLyS: Human IgG1	Autoimmune	Human Genome Sciences	2011 (US,EU)
Yervoy: Ipilimumab	CTLA-4: Human IgG1	Cancer	Bristol-Myers Squibb	2011 (US,EU)

Adapted from \*Chames et al. (2009), Ecker et al. (2015), FDA

**Table 1-3: Capital investment costs for antibody-based facilities using mammalian cells\***

Manufacturing facility	Date facility completed	Capital Investment (\$ US M)	Area (sq. feet)	Production bioreactor capacity		
				Number	Size (L)	Total (L)
Boehringer ingelheim expansion —Biberach, Germany	2003	315	—	6	15,000	90,000
Lonza Biologics expansion—Portsmouth, NH, USA	2004	207	270,000	3	20,000	60,000
Amgen—BioNext, West Greenwich, RI, USA	2005	500	500,000	9	20,000	180,000
Genentech expansion—Oceanside, CA, USA	2005	380	470,000	6	15,000	90,000
Imclone expansion—Branchburg BB50, NJ, USA	2005	260	250,000	9	11,000	99,000
Biogen Idec—Hillerød, Denmark	2007	350	366,000	6	15,000	90,000
Lonza biologics—Tuas, Singapore	2009	250	—	4	20,000	80,000
Genentech expansion—Vacaville, CA, USA	2009	600	380,000	8	25,000	200,000
Bristol-Myers Squibb – Devens, MA, USA	2011	750	—	6	20,000	120,000
Pfizer Biotech Campus, Grange Castle, Ireland	2011	1800	—	6	12,500	75,000
MedImmune, Frederick, MD, USA	2011	600	337,000	4	12,500	50,000

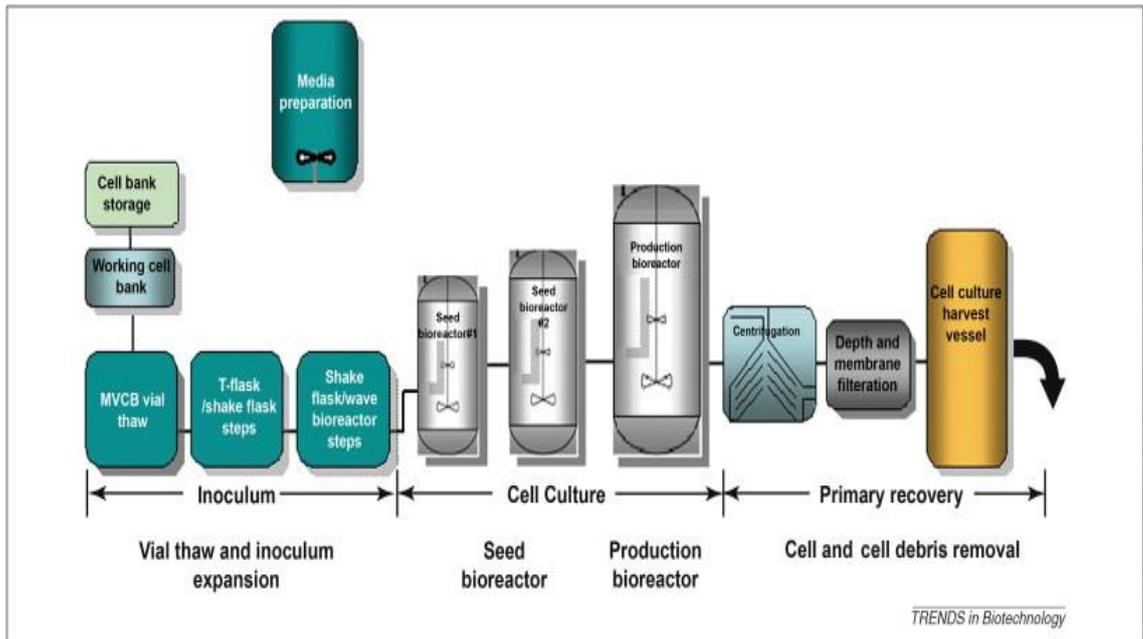
(Adapted from \*Pollock (2013))

## **1.2 Antibody manufacture**

The immune's system B-lymphocytes produce monoclonal antibodies in response to foreign proteins named antigens. The primary function of antibodies is recognizing and attacking foreign bodies like bacteria and viruses, which invade the organism. The history of therapeutic antibodies began in 1986 when the FDA approved the first of a generation of murine derived antibodies, OKT3™. The use of murine derived antibodies has its limitations, such as the production of anti-murine antibodies (HAMA) by the human immune system. Further development gave rise to ReoPro™, the first chimeric antibody (only the light chains of the antibody were of murine origin) in 1993 by Johnson & Johnson. In this case the human immune response was not as strong as with OKT3™. This immunogenicity had to be reduced, therefore scientists tried to find a way to eliminate the murine component. As a consequence, in 1998, Genentech introduced on the market the first humanized antibody, Herceptin™, developed to fight against breast cancer. Then in 2002, FDA approved the first human antibody produced by phage display technology, Humira™ (AbbVie Inc). Most of the approved mAbs in current use are either chimeric or humanized (Sommerfeld and Strube, 2005; Chames et al., 2009).

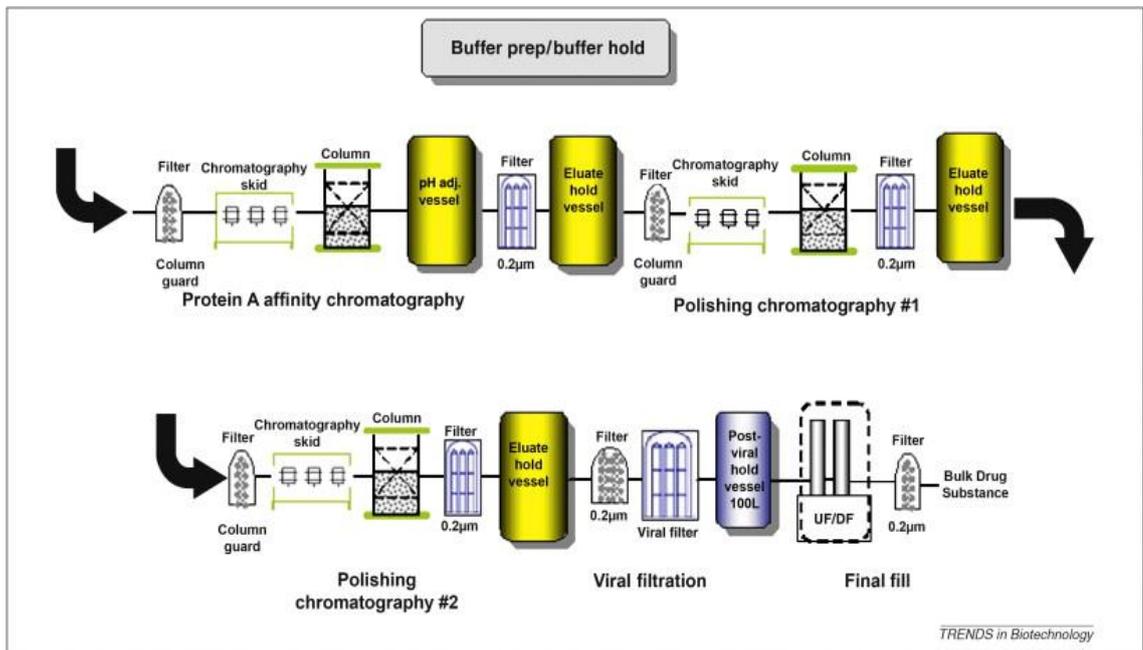
### **1.2.1 Antibody platform process**

Platform processes are defined as a series of manufacturing operations that can be relevant to more than one product, with the aim of reducing manufacturing investments and eliminating process re-design for each new product. Antibody manufacture adopts a platform approach with standard unit operations used for the downstream process including clarification using centrifugation as well as depth and membrane filtration, antibody capture through Protein A affinity chromatography, polishing, virus removal and formulation (Curling, 2004). Platform processes give a general approach to antibody production, which significantly lowers the development time (Farid, 2009). An antibody platform often used for monoclonal antibodies is presented below.



**Figure 1-1 : Typical upstream process for monoclonal antibodies**

(Adapted from Shukla and Thommes (2010))



**Figure 1-2: Typical downstream process for monoclonal antibodies**

(Adapted from Shukla and Thommes (2010))

In the upstream process cells are grown stepwise. Firstly, the cells are frozen in a master cell bank and various working cell banks. Then, the cells are cultured in shake flasks and grown in gradually larger volumes for a few weeks. This will then constitute the inoculum for large-scale fermentation (10,000-20,000 L). Once the cells are added to the large-scale bioreactors, they are cultured for approximately 2 weeks, predominately as fed-batch processes. For large-scale production of mAbs a typical bioreactor train is shown in Fig. 1-1 (Sommerfeld and Strube, 2005).

The harvest material is first clarified by either centrifugation or filtration, widely used current techniques for harvest operations. Centrifugation may not give an acceptable degree of solid removal in order to be able to load the centrate directly onto the first chromatography capture step. A filtration step using depth filters may also additionally be required. Depth filters are capable of removing host cell protein (HCP), any potential cell debris and other impurities. It has also been shown to reduce the level of turbidity seen in Protein A eluate of a monoclonal antibody process (Yigzaw et al., 2006). Protein A Chromatography involving a low pH elution step captures and purifies monoclonal antibodies. This step also acts as a viral inactivation step. Two polishing steps such as cation- (CEX) and anion- (AEX) exchange chromatography are usually employed in order to meet the purity requirement imposed by the regulatory authorities. A virus filtration step is necessary to provide additional assurance that the virus is removed and the product is safe, followed by a final ultrafiltration/diafiltration step which helps formulate and concentrate the product (Fig. 1-2). Overall purification yields range from 70-80%. Different companies use similar platform processes as it can be seen in Table 1-4. (Shukla et al., 2007; Kelley, 2009; Shukla and Thommes, 2010; Zhang et al., 2014).

Looking at several existing manufacturing processes employed by companies for the production of marketed monoclonal antibodies (Table 1-4) it can be observed that clarification (achieved by centrifugation or microfiltration) is the common first step in these processes. In the majority of cases, Protein A affinity chromatography is used as a capturing step which also incorporates a viral inactivation step due to its low pH elution. Several ion exchange steps follow protein A. These polishing chromatographic steps are used for the removal of DNA fragments, HCPs, potential

leaches from Protein A and other impurities. If the required purity is not attained, extra chromatographic steps such as hydrophobic interaction or size exclusion may be considered (Sommerfeld and Strube, 2005).

**Table 1-4: Analysis of different production processes for marketed antibody-based therapeutics\***

	<b>Herceptin</b>	<b>Rituxan</b>	<b>MabCampath</b>	<b>Synagis</b>	<b>Remicade</b>	<b>Simulect</b>
Cell removal	1	1	1	1	1	1
Affinity chromatography	2	2	2		2	2
Virus inactivation	3	3	3	4	3	3
Cation exchange	4	5	4	2	4	5
Anion exchange	5	4		3, 6	6, 7	4
Hydrophobic interaction	6					
Size exclusion chromatography			5	8		
Virus clearance		6	6	5, 7	5	6
Sterile filtration	7	7	7	9	8	7

\*Adapted from Sommerfeld and Strube (2005)

### **1.2.2 Quality by Design**

The concept of Quality by Design (QbD) for biopharmaceutical drug development was initiated by the Food and Drug Administration (FDA) and involves a better understanding of the product and its manufacturing process, aiming to build product quality into the process by design, as opposed to testing for it at the end. QbD comprises of three main components: process knowledge involving a good understanding of the impact process inputs have on process performance, the identification of potential critical quality attributes (CQA) and understanding the relationship between the manufacturing process and these CQA as well as the link between CQA and product's clinical properties. A CQA can be defined as a quality attribute of the product that has the ability to influence safety and efficacy (Abu-Absi et al., 2010; Eon-Duval et al., 2012a). The Quality by Design approach for characterising mammalian cell culture processes has been widely used (Abu-Absi et al., 2010; Horvath et al., 2010; Rouiller et al., 2012).

### 1.2.3 Increases in mAb titres

The bioreactor concentration of recombinant products, cultured in mammalian cells has seen an increase of over 20-fold within the last two decades. The main reasons for these advances have been improvements in the capability to isolate high producer cell lines through gene amplification and cell isolation as well as the development of improved fed-batch protocols (Butler and Meneses-Acosta, 2012). Further factors contributing to the increase in production yield are optimisation of culture media, addition of small molecule enhancers and improvement to process control such as dissolved oxygen, pH, temperature and media osmolality (Bai et al., 2011). These advances are able to increase cell productivity, extend culture duration and maintain a high viability for longer (Lu et al., 2013). Bai et al. (2011) showed that a combination of iron and citrate added to chemically defined media increased mAb production in CHO cells by 30-40 %, without impacting on product quality. Modelling approaches, such as multivariate data analysis helped improve the sector's knowledge with regards to culture conditions in order to optimize cell growth and improve product titre (Elvin et al., 2013). Nowadays, mAb concentration of 3-5 g L<sup>-1</sup> are commonly achieved and titres up to 10-13 g L<sup>-1</sup> in fed-batch processes are reported (Kelley, 2009; Li et al., 2010). Antibody titres as high as 17 and 25 g L<sup>-1</sup> have also been attained. Titres of 17 g L<sup>-1</sup> have been reported in concentrated fed-batch cultures while titres of 25 g L<sup>-1</sup> were reported in perfusion cultures. Both systems are based on the alternative tangential flow (ATF) system (Shevitz et al., 2011). Antibody titres of 25 g L<sup>-1</sup> were achieved in PER.C6 human cells using a continuous cell culture process operated in "concentrated fed-batch" mode (eXtreme-Density, XD). This was performed using a modified ATF perfusion system to retain the cells and the product in the bioreactor (Schiermer et al., 2010; Chon and Zarbis-Papastoitsis, 2011). Improvements in mAb titres are essential in order to assist high drug dosage, increased demands and be able to control production costs (Robert et al., 2009).

Improvements in cell culture processes have resulted in higher mAb titres by increasing culture duration, cell densities and specific cell productivity (Brodsky et al., 2012). This involves bioreactors operating at high cell densities (>1 × 10<sup>7</sup> cells

mL<sup>-1</sup>), leading to higher levels of host cell proteins, nucleic acids, media and feed components (Westoby et al., 2011). XD cultures exhibit very high cell densities ( $> 150 \times 10^6$  viable cells mL<sup>-1</sup>) compared to conventional levels of  $10\text{-}20 \times 10^6$  cells mL<sup>-1</sup>, in fed-batch cultures. These high levels of cell densities put increased pressure on harvesting techniques (both centrifugation and filtration) due to their high solids content ( $\leq 40\%$ ). Changes in upstream processes require optimisation of downstream process methods (Tscheliessnig et al., 2013).

Processes featuring higher titres can pose facility fit challenges in biopharmaceutical purifications suites (Brodsky et al., 2012; Yang et al., 2014). The large volumes of buffers, chromatography resins and large filter areas needed to cope with higher titres, are generally incompatible with current manufacturing facilities (Aldington and Bonnerjea, 2007; Kelley, 2009; Stonier et al., 2012). Mismatch in equipment sizes can cause bottlenecks and together with process fluctuations upon scale-up leads to discarding expensive product. Yang et al. (2014) explains that large scale manufacturing facilities which have fixed equipment, find it harder to adopt debottlenecking strategies that involve equipment changes in response to fit issues derived from higher titres. They found that titre, CEX and AEX eluate volumes were the most significant factors impacting unexpected mass loss and they propose three debottlenecking solutions. Using larger volume pool tanks for AEX, CEX and viral inactivation steps, reducing the eluate volumes of the CEX and AEX steps as well as using higher capacity resins for these steps could minimise product loss.

These high mAb concentration feed volumes are putting increased pressures on legacy facilities with downstream equipment that was designed for a much lower amount of mAb (Stonier et al., 2012; Yang et al., 2014; Tao et al., 2014). These pressures arise from the increased level of solids affecting harvesting operations and the higher amount of impurities (higher level of total HCPs and problematic HCPs) affecting purification steps. The harvest and purification equipment reach their limits in terms of capacity, which results in higher material consumption, processing times and cost. At low mAb titres, the cost of upstream manufacturing is higher than downstream but higher titres shift the cost from USP to DSP (Chon and Zarbis-Papastoitsis, 2011; Levy et al., 2014). The downstream costs for CHO processes

represent between 50-80 % of the total cost of manufacturing a recombinant protein (Lowe, 2001; Guiochon and Beaver, 2011). In order to minimise these cost and to ensure the DSP is able to cope with the upstream feed, improvements need to be made to current technologies as well as looking into cheaper, more efficient technologies. Alternative steps should be able to either purify the proteins individually or minimise the burden on the current purification steps through partial purification (Butler and Meneses-Acosta, 2012; Saraswat et al., 2013).

#### **1.2.4 Future trends of antibody platform**

The upstream process platform for monoclonal antibodies production has considerably progressed in terms of the cell line development and bioreactor process that resulted in high product titres. Even though the current process platform is safely and widely used in industry, a number of drivers will lead to further development in the next decade. Due to the recent development in achieving high titres, the necessity to build biopharmaceutical facilities with high volume bioreactors (eg. 20,000L) will decrease. Producing lower volumes will lead to higher use of disposables in manufacturing and will emphasize the need for facility flexibility (Shukla and Thommes, 2010).

New cell culture technologies are being evaluated to make the processes even more robust and to reduce the operation costs. Some of these new technologies include the development of new on-line process monitoring and control systems as well as the development of high-throughput cell culture systems. Cell culture monitoring using accurate in situ sensors for measuring relevant parameters could improve the development and optimisation of cell culture processes. Stainless steel tanks are the current choice for laboratory and pilot scale process development as well as for the large-scale manufacture of antibodies. The capital investment and maintenance cost for fixed plant equipment is very high and the validation requirements in terms of cleaning and sterility are high as well. These issues can be overcome by the use of disposable bioreactors, which have become extensively used in mammalian cell bioprocesses. They have the advantages to increase plant capacity and flexibility by reducing the turnaround time, easier and more rapid to implement

design changes and smaller footprint (Li et al., 2010; Butler and Meneses-Acosta, 2012).

Product recovery for centrifugation can be low due to an increased need of desludging (particularly in large-scale continuous centrifugation). Clarification post centrifugation can also be affected due to shear forces cell disruption. Depth filters alone are unable to handle high-solids feedstreams and are usually used in combination with centrifugation. Tangential flow filtration (TFF) is able to handle high solids but the yield might be poor. Schirmer et al. (2010) presented an alternative method to centrifugation, which in combination with filtration would benefit the clarification of high-density cell harvests. The enhanced cell settling method (ECS) involves the addition of weak IEX matrices to the cell culture harvest, which enhances cell settling. They demonstrated that this method greatly reduces HCP and DNA levels in partially clarified harvests. Westoby et al. (2011) showed how a reduction in cell culture fluid pH to 4.7-5 induced flocculation and precipitation of impurities which increased the average particle-size. This method enhanced impurity removal and improved tangential flow microfiltration throughput and filter capacity.

Most of the manufacturing processes use Protein A chromatography as the capturing step, but there are some potential disadvantages associated with Protein A chromatography such as high resin cost, potential ligand leaches and sensitivity to product residence time (Chon and Zarbis-Papastoitsis, 2011; Saraswat et al., 2013). Alternatives to Protein A for capture of mAbs include batch chromatography (cation-exchange chromatography (CEX) (used in the process of manufacturing HUMIRA and Synagis), mixed-mode chromatography and continuous chromatography. Cation-exchange processes used to have low capacity (20-30g/L) and in the past they were not able to deal with high cell culture titres. Most recently processes have been developed with higher CEX capacities, of 100 g/L, for monoclonal antibodies, demonstrated using real feeds as well as model systems. These newly improved chromatographic media have the ability to achieve high capacity at only 1/5 of Protein A cost, making them a more appealing alternative to affinity chromatography (Gagnon, 2010; Chon and Zarbis-Papastoitsis, 2011). Tao et al. (2014) showed the

feasibility of using CEX for direct capture of mAbs from high titre cell culture harvest. Another alternative for the capturing step is mixed-mode chromatography (anion exchange and hydrophobic interaction) which can be optimised by design of experiments and can present comparable yields to those seen in Protein A, but at a much lower cost (Touelle et al., 2011). Pezzini et al. (2011) used mixed-mode chromatography for capture of mAbs from CHO cell culture supernatants. They also evaluated four types of mixed-mode resins, in their ability to minimise HCP levels. Semi-continuous chromatography (e.g. periodic counter current (PCC)) has been successfully applied to the capture of mAbs from cell culture supernatant. Mahajan et al. (2012) showed the application of three small-scale columns operated using the PCC principles for the purification of low and high mAb titres. It also highlighted a 40 % reducing in the cost of resin, buffer and processing time when the multi-column chromatography was compared to affinity chromatography. Pollock et al. (2013) evaluated the application of semi-continuous 3-column and 4-column PCC chromatography system for the capture of mAbs and showed that the use of semi-continuous chromatography can reduce manufacturing costs of early clinical phase material.

Non-chromatographic purification methods such as membrane chromatography or selective precipitation are likely to arise (Shukla et al., 2007). Membrane chromatography uses filtration membranes with ligands immobilised to the inner pore surface resulting in a selective adsorption of molecules, separating them based on their chemical behaviour (Frohlich et al., 2012). Membrane chromatography has been used in both flow through and bind-and-eluate mode in industrial applications (Hirai et al., 2009; Liu et al., 2010). Precipitation can be used as an impurity removal step prior to harvest (using caprylic acid) (Brodsky et al., 2012) or capture by CEX (using PEG) (Lain et al., 2010). Brodsky et al. (2012) showed the use of caprylic acid precipitation being implemented in a bioreactor prior to harvest as well as being used as an alternative to polishing chromatography, following Protein A affinity chromatography.

### 1.3 Small scale cell culture

The relationship between culture conditions and process outcomes is very complex. A multi-factorial experimentation strategy can be used to analyse relationships and provide the highest amount of data in the shortest amount of time (Montgomery, 2009). It is difficult to apply this strategy to bench-top or large-scale bioreactors due to high resource requirements as well as high capital equipment and infrastructure costs (Lewis et al., 2010). The need to perform large numbers of experiments under controlled, bioreactor conditions has resulted in the development and implementation of high-throughput, small-scale systems for process development (Legmann et al., 2009; Lewis et al., 2010; Jones et al., 2015).

Among the small-scale devices investigated are shake-flasks, microtitre plates and stirred bioreactors (Kumar et al., 2004; Betts and Baganz, 2006; Duetz, 2007). Fed-batch cultivations are not easily conducted on a routine basis and under controlled conditions making shake flasks less favourable for bioprocess optimisation. To overcome these limitations new small-scale and high throughput (HT) systems are now being commercialized. These are better able to mimic the performance of stirred tank bioreactors and provide sufficient data. HT technologies facilitate faster timelines, shorter development times while fewer resources are required. (Amanullah et al., 2010). Various designs of HT micro-scale bioreactors have been reported in literature, starting from simpler standard plate with integrated sensors measuring pH and oxygen levels (microtitre plate based bioreactors, e.g. Simcell), to more advanced cell culture systems with better control capabilities (stirred mini bioreactors, e.g. Micro 24<sup>®</sup> MicroReactor (Pall Life Sciences, Port Washington, NY) and the ambr<sup>®</sup> (Sartorius Stedim, Royston, UK). To be able to support the HT micro-scale bioreactors, HT analytical techniques such as the Gyrolab workstation (Gyros, Uppsala, Sweden) (quantification of HCP concentration) have also been implemented in industry.

### **1.3.1 Traditional small-scale culture systems**

#### **1.3.2 Shaken vessels**

##### **1.3.2.1 Shake flasks**

Shake flasks are widely used and have the advantage of low price and being easy to handle. Shake flasks are manufactured in different sizes with capacity ranging from 25 mL to 5L, made of plastic or borosilicate glass and equipped with or without baffles. They are usually operated on orbital shaking devices at specific shaking frequencies while only the temperature is controlled. The mixing and mass transfer within the flasks is achieved due to orbital shaking (Büchs et al., 2000). Flasks not containing baffles should be operated in a way in which bubbles are not formed and this will provide good gas-liquid mass transfer conditions (Büchs, 2001). pH is kept within a reasonable range with the help of buffers and with the use of recent optical sensor system. Dissolved O<sub>2</sub> can also be accurately measured (Tolosa et al., 2002; Witmann et al., 2003). pH can be measured by using invasive pH probes. The dependency of shake flasks on orbital shaking for agitation and on surface aeration makes reduced oxygen transfer a major limitation in comparison with stirred tank reactors (Bareither and Pollard, 2011). Baffled shake flasks with high agitation and smaller working volume can be used to maximize oxygen transfer (Lotter and Büchs, 2004). The use of shake flasks for process development is limited not only because of oxygen transfer issues but also because of little similarity to stirred tank bioreactors, difficulty to control DO, pH and maintain feed capability (Bareither and Pollard, 2011).

##### **1.3.2.2 Spin tubes**

To increase the throughput and reduce the size of shake flasks, De Jesus et al. (2004) developed a scale-down system named “Tube Spin”. Centrifuge tubes of 50 mL are used for culturing, set on a rotational shaker in the incubator. In situ control and measurement of DO and pH cannot be performed, but the system can still be used for large screening experiments, primarily for cell line selection and media optimisation.

Vented caps can be used for this system in order to allow gas exchange via the headspace. Tube Spin can be connected to an automated liquid handling system to increase its capabilities.

### **1.3.2.3 Microtitre plates**

Due to their increased throughput, microtitre plates (MTP) are increasingly replacing shake flasks (Silk et al., 2010). The pharmaceutical industry invested significantly in the development of high throughput screening (HTS), in order to accelerate drug discovery. Microtitre plates can be manufactured in a wide range of sizes (6-1,536 wells) however only 24, 48 and 96 well formats are commonly used in the development of bioprocesses. In strain selection process for primary screening the 96 well plates is normally used (Bareither and Pollard, 2011). Common volumes used in microtitre plate vary from 0.025 mL to 5 mL therefore the use of these provide at least 50-fold decrease in medium needed compared to shake flasks as well as a cost reduction (Silk et al., 2010). This process can be automated with the help of robotics with modern pipetting and dispensing systems. In this way, numerous samples can be managed in a short period of time. Being able to perform parallel cultivation in microwell plates with a reduction in scale and lower labour costs demonstrates that new biopharmaceuticals produced will be able to get to the market more quickly (Silk et al., 2010).

Many studies regarding mixing and oxygen mass transfer rate (OTR) in MTPs have been published (Lye et al., 2003; Harms et al., 2006), which shows that OTR is sufficient to support the needs of suspension cell culture (Chen et al., 2009). Microtitre plates face challenges related to liquid evaporation rates while keeping acceptable aeration and gas exchange rates (Chen et al., 2009) and contamination risks produced by aerosol formation at high shaking rates (Kumar et al., 2004). To tackle the online monitoring problem of microtitre plates, non-invasive fluorescent technology is used by the incorporation of fluorescence patch sensors into the base of each well for monitoring and measurement of OD, pH and DO (Chen et al., 2009; Bareither and Pollard, 2011). Silk et al (2010) presented a technique for the successful fed-batch culture of mammalian cells (GS-CHO cell line) in shaken 24-

standard round well plates with growth rates, antibody productivities and viabilities that are the same as those in standard shaken flasks. This work demonstrated the possibility of carrying out fed-batch mammalian cell cultures in shaken micro well formats.

#### **1.3.2.4 Micro 24 bioreactor**

The Micro 24 is a miniaturized bioreactor system used as a scale-down model for cell culture process development. It uses a specialized 24 deep-well plate as shown in Fig 1-3. Each individual well can be controlled like an independent bioreactor. In this way, throughput can be increased whilst maintaining data quality and quantity. Each well contains non-invasive sensors for temperature, pH and DO, as well as a 0.2  $\mu\text{m}$  sparge membrane for gas blending using air and a thermal heat conductor. To seal each well, vent caps are used.



**Figure 1-3: Micro 24 bioreactor**

Chen et al. (2009) first showed the assessment of the Micro 24 for process development of mammalian cell culture. They compared the performance of the miniaturized bioreactor (5 mL w/v) with that of a 3L (2 L w/v) bench-scale bioreactor using fed-batch culture of CHO cells. DO, pH and temperature were controlled within each well and agitation was controlled at plate level. In term of gases, N<sub>2</sub>, clean air and CO<sub>2</sub> were monitored and controlled. These experiments were carried out to examine the reproducibility between all 24 wells and scalability between M24 and the 2 L bioreactor. Similar results have been shown for the two systems in terms of percent cell viability, culture doubling time as well as the integrated viable cell counts. Ammonia, lactate, glucose and glutamate profiles were

also similar when comparing the M24 and 2L bioreactor cultures. This work shows that M24 is a good scale-down model for cell culture applications but no engineering basis for the M24 – 2L bioreactor comparison was presented.

### **1.3.3 Stirred systems**

Stirred systems are generally better than most of the static or shaken systems regarding online monitoring and control. Stirred systems give a high degree of freedom to increase mass transfer and mixing by increasing gassing rate and stirred speed as well as offering a consistent environment (Kumar et al., 2004).

#### **1.3.3.1 Spinner-flasks**

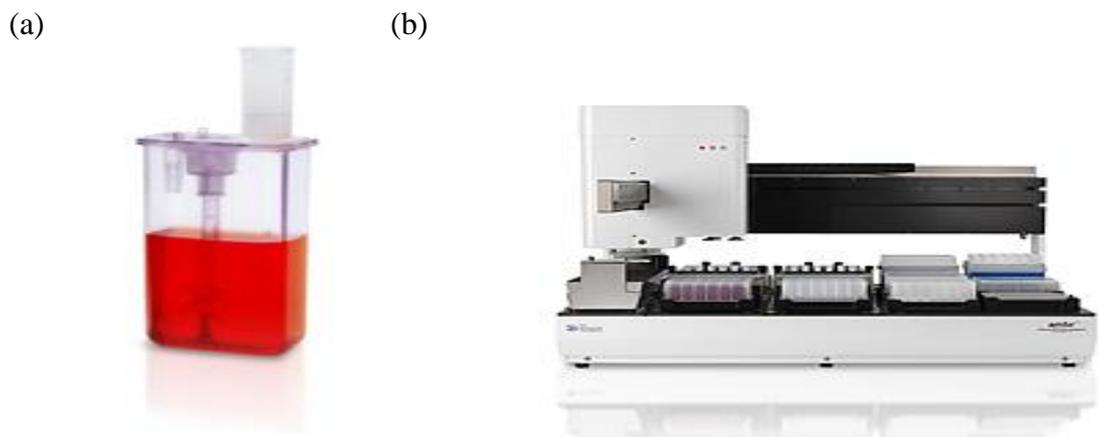
Spinner-flasks were the first stirred bioreactors designed for animal cell growth (Wang et al., 2002). A standard spinner-flask has a stirred shaft and side arms with screw caps. The medium and inoculum can be added through these arms as well as different types of probes such as pH and temperature, for gassing the flask with O<sub>2</sub> and CO<sub>2</sub> (Kumar et al., 2004, Yeatts and Fisher, 2011). The maximum culture volume which can be used is dependent on the cell types, how often the culture media is changed and how the culture conditions vary throughout the culture cycle. Online measurement of cells oxygen uptake rate (OUR) and pH control can be achieved. CHO suspension cell cultures can be easily cultivated in spinner-flasks (Kumar et al., 2004).

#### **1.3.3.2 Stirred minibioreactors**

Stirred minibioreactors are a miniature version of conventional stirred bioreactors. These bioreactors are fabricated from borosilicate glass, polycarbonate or stainless steel having the top head-plate made of PEEK (polyetheretherketone) or polycarbonate. These are equipped with various ports for probes such as temperature, pH and dissolved O<sub>2</sub>. Common volumes used vary from 50 mL to 300 mL. Stirred minibioreactors have comparable performance to traditional bioreactors and even if these are small, critical operating parameters such as temperature, pH and dissolved

O<sub>2</sub> can still be monitored and controlled at the required levels. These are ideal for research studies or later stage cell culture process development, as small amounts of medium are required in order to run long-term fed-batch operations. Some of the disadvantages associated with these are their limitations in high-throughput screening and their high cost in comparison with other small scale devices as well as a high degree of manual operations such as cleaning and sterilization (Kumar et al., 2004).

### 1.3.3.3 ambr system (advanced micro scale bioreactor)



**Figure 1-4: Illustration of (a) ambr microscale bioreactor vessel (b) automated ambr workstation**

The ambr is an advanced micro scale bioreactor, which attempts to mimic the performance of large scale bioreactors at a 10-15 mL microscale, using disposable reactor cartridges. The selection of improved cell lines can be achieved faster due to the system's ability to enable fast evaluation of various bioreactor cultures while providing significant savings in terms of materials and labour. The ambr system has 3 key components: the ambr bioreactor vessel (Fig. 1-4 (a)), which provides the scale down mimic, the automated ambr workstation (24 or 48) (Fig. 1-4 (b)), which allows parallel bioreactor cultures and saves user time and the ambr software that helps process the data. The ambr bioreactor vessel is composed of non-invasive sensors for measuring and controlling DO and pH, placed at the bottom of the bioreactor. In contrast to the micro 24 system (Section 1.3.2.4) it has a marine impeller just like the one used for large bioreactors, a sparge tube for the gas supply with an in-line filter and a vessel cap designed for reliable robotic removal and replacement. Automation

is achieved for culture set-up, inoculation, liquid additions (media, cells, antifoam, feeds), sampling, therefore the risk of cross contamination is minimized.

Lewis et al. (2010) evaluated the capability of the ambr 24 system to replicate the characteristics of classical bioreactors at micro-scale by comparing it to a 7L bench-top bioreactor. It has been shown that viability profiles and antibody titres obtained in the ambr 24 system under fed-batch conditions are in good correlation with those seen in the 7L bioreactor. Batch cultures were carried out simultaneously, in all 24 vessels under identical conditions in order to demonstrate reproducible growth profiles across all 24 bioreactors. It was demonstrated that the ambr system shows good vessel-to-vessel consistency in terms of viable cell number as well as consistent antibody titres between individual bioreactors and between the two culture stations within the system. Similar work performed by Hsu et al. (2012) and Moses et al. (2012) showed good comparability between ambr 24 and 2L, 3L bench-top bioreactors, respectively, in terms of online controls, culture performance, metabolites and product quality. Nienow et al. (2013) showed good comparability between the ambr system and 5L bioreactors as well as investigating the physical characteristics of this microscale bioreactor system. Rameez et al. (2014) also showed highly reproducible results between the ambr 24 and 3L, 15L and 200 L stirred tank bioreactors in terms of cell growth, process capabilities and product titre.

#### **1.3.3.4 Microfluidic “SimCell” bioreactor**

Seahorse Bioscience Inc. (Billerica, MA) has developed the SimCell system, which is a high-throughput micro-bioreactor scale-down model based on microfluidic technologies (Kim et al., 2011; Kim et al., 2012). This system is a fully integrated robotics platform that was created to reproduce the required conditions within suspension culture bioreactors. The platform contains micro-bioreactor arrays, each composed of 6 micro-bioreactors. These bioreactors are secured in a cassette-type system containing transparent, biocompatible, membranes that allow gas transfer, used to reach up to 20 million cells mL<sup>-1</sup> (Heath and Kiss, 2007; Legmann et al., 2009). Each micro-bioreactor has a volume smaller than 800 µL. Oxygen and carbon dioxide transfer is possible in each micro-bioreactor due to the gas permeable

membrane. Losses due to evaporation are minimized due to their controlled design. Agitation of multiple cassettes can be achieved by rotational agitation (20 rpm) inside each incubator. Computational fluid dynamics (CFD) confirmed that this mimics the expected shear rate of conventional stirred tank bioreactors (Bareither and Pollard, 2011). The SimCell™ System is comprised of five incubators, each of them able to hold maximum 42 micro bioreactors (MBs) with 1,260 experiments in total. At the incubator level, parameters such as gas composition for DO and pCO<sub>2</sub>, temperature and relative humidity can be controlled. Fluorescence detection can be used for measuring pH and dissolved oxygen in the micro-bioreactors (Amanullah et al., 2010). The “SimCell” system is able to achieve high throughputs (more than 1000 concurrent experiments) and data generation rates due to a fully automated robotics system capable of sampling, controlling temperature, pH and monitoring dO<sub>2</sub>, cell mass by optical density within each of the six microbioreactors on each plate (Heath and Kiss, 2007).

Amanullah et al. (2010) presents the application of SimCell™ micro-bioreactor for the fed-batch culture of GS-CHO transfectant expressing a model IgG4 monoclonal antibody. For examining process scalability and reproducibility in a 250 mL shake flask, 3 and 100 L bioreactors, 114 parallel MBs were used. The performance of the MBAs, including viable cell density, MBA protein titer and metabolite profiles were compared to those obtained in the shake flask, 3 and 100 L bioreactor cultures. Titre profiles, glucose concentration profiles and lactate concentration were comparable to the scale-up versions and within ±20 % of historical data. The results show that the SimCell platform operated in fed-batch mode with pH, glucose and DO control is able to successfully reproduce both shake flask and bioreactor cultures and sustain viable cell concentrations up to  $12 \times 10^6$  cells mL<sup>-1</sup>.

#### **1.4 Impurities in biopharmaceutical manufacturing**

The impurities present during the manufacturing process of therapeutic proteins such as mAbs can be divided into two main categories: product and process related impurities. Product-related impurities could comprise of unwanted molecular

variants of the product such as oxidized forms, precursors, aggregates or degraded products. Process-related impurities arise from the manufacturing process itself (either upstream or downstream). Process-related impurities include host cell proteins (HCPs), DNA, chemical additives (inducers, antibiotics, protease inhibitors) as well as impurities from the downstream process (e.g. leachables such as Protein A, plastics, heavy metals) (Shukla et al., 2008; Tscheliessnig et al., 2013).

#### **1.4.1 Host cell proteins (HCPs)**

HCPs are a unique and complex group of impurities that can be present in the supernatant due to secretion of the host cell or released through cell breakage and reduced viability of the cells. They account for a large subgroup of process-related impurities (associated with the type of process used and not the product itself). It has been reported that their presence in the final recombinant mAb product can raise safety concerns, as particular HCP species could cause adverse clinical effects in humans, even when low levels are present. For this reason it is required to closely monitor them and demonstrate that the downstream manufacturing process is able to reduce HCPs to acceptable low levels as detected by a sensitive analytical method (FDA, 1997). Typical target values of impurities in the final recombinant mAb products are <100 ppm of host cell proteins, <5% high molecular weight immunogenic aggregates and 10 ng/dose of DNA (Wang et al., 2009; Chon and Zarbis-Papastoitsis, 2011; Levy et al., 2014; Reisinger et al., 2014).

HCPs are complex in their structure and composition, displaying a wide range of properties. A group of HCPs resulting from one process can be very diverse compared to another process, presenting different structure, isoelectric points, molecular mass and hydrophobicity. They can pose significant challenges to the downstream process, due to the varying amount, composition and characteristics during a particular cell culture process (Gronemeyer et al., 2014). The protein of interest, the host organism, subcellular localization of expression, cell culture and harvest conditions have been shown to influence the abundance and composition of HCPs present in the harvest material (Shukla et al., 2008; Wang et al., 2009; Jin et al., 2010; Gutierrez et al., 2012; Tarrant et al., 2012). Jin et al. (2010) explored the

effect of several upstream process parameters on HCP profile and found culture viability had the most significant impact on the HCP profile. A similar result was also published by Tait et al. (2012) which found that most of the HCPs were present in the supernatant, originating through lysis or cell breakage associated with a decrease in viability. Hogwood et al. (2013) showed that the choice of depth filters during primary recovery modified the HCP profile and HCP concentration in harvested cell culture fluid. Schirmer et al. (2010) reported a reduction in harvest HCP levels of 60 % when the enhanced cell settling method (ECS) was used while Shukla et al. (2008) reported 2.3 logs of HCP clearance (LRV) over Protein A step. This figure dropped to 1.4 when the “worst-case” strategy was applied. The “worst-case” strategy involved operating steps at a combination of operating parameters that give the poorest clearance of HCPs over that step.

A sub-group of these HCPs are known as problematic and prove difficult to remove during purification due to their association with the mAb or with the chromatographic matrix (Dorai et al., 2011; Aboulaich et al., 2014; Levy et al., 2014; Valente et al., 2015). Varying combinations of USP factors that have the potential to alter the impurity profile which is then carried on to the downstream process can make clearance of these difficult to remove impurities more complicated. Wu (2013) showed that the USP composition of HCPs can impact impurity clearance during purification. Individual problematic HCPs should be identified early in process development (USP) and their clearance monitored during downstream. During process development it is essential to understand the impact of the USP and DSP process on the HCP profile. This could help reduce the cost of goods by minimizing the burden of HCPs on the DSP or through the implementation of alternative, more cost effective methods. Ensuring a minimal HCP content in the supernatant could reduce the risk of any residual immunogenic HCPs to be found in the final drug product (Hogwood et al., 2014).

#### **1.4.2 Analytical tools for detection, monitoring and quantitation of HCPs**

Highly sensitive and appropriate analytical methods are required for measuring HCPs at various stages within the manufacturing process, in order to support process

development, validation and to ensure regulatory approval. The ideal method (or combination of methods) should be able to produce quantitative results, to identify all HCP species present in a sample, to provide high-throughput measurements and have a short processing time (Tscheliessnig et al., 2013). The current methods currently used to monitor, measure and identify HCPs are split into two categories: immuno-specific and non-specific. Immuno-specific methods identify HCPs by using polyclonal anti-HCP antibodies produced by immunization of a production animal (rabbits or goats) using supernatant or partially purified material from a relevant null cell line culture. A null cell line refers to a production cell line that does not have the product coding gene (Jin et al., 2010; Tscheliessnig et al., 2013). Nearly all biopharmaceutical companies use antibody reagents that were generated especially for their cell lines. Immuno-specific methods include ELISA, western blot and slot blot while non-specific methods include gel electrophoresis and mass spectrometry. Commonly used methods are described below.

#### **1.4.2.1 ELISA**

Enzyme-linked immunosorbent assay (ELISA) is an immuno-specific method and is the current gold-standard approach used to detect and measure total HCP concentrations during bioprocessing manufacturing and in the final drug substance. These assays are highly selective; provide high sensitivity and high-throughput mode operation. While ELISA can provide a good evaluation of the total level of HCPs in a given sample, it does not give any information on the specific HCP species that are found within the whole population. This prevents determining the risk associated with certain problematic HCPs (Jin et al., 2010; Bracewell and Smales, 2013; Hogwood et al., 2014; Thompson et al., 2014; Zhu-Simoni et al., 2014; Bracewell et al., 2015). ELISA cannot be used to identify qualitative changes in HCP population (Reisinger et al., 2014). Another drawback with this assay is the fact that in any anti-HCP antibody pool, there are no antibodies that are able to identify every HCP species that might be present. There is the likelihood that very weak immunogenic or non-immunogenic species will not be identified (Bracewell et al., 2015). The sensitivity of this method relies on the antibody reagent used and generic HCP assay

kits might not be able to identify a wide range of HCPs, resulting from different cell line. Cell-line specific assays would provide improved sensitivity. ELISA has been an invaluable tool for measuring and monitoring HCP concentrations during process development as well as in the final drug product and it will continue to be widely used by the biopharmaceutical industry to monitor and control HCPs. Another method for quantifying HCP levels is fourier transform mid infrared spectroscopy (FT-MIR). This method has the advantage of being able to measure the HCP concentration directly in the bioreactor or in a bypass. Capito et al. (2012) used FT-MIR to quantify HCPs in a CHO cell culture fluid producing mAbs, after treatment with different polyelectrolytes for semi-selective clarification. The results were comparable to those obtained by ELISA.

#### **1.4.2.2 Gel electrophoresis**

Since ELISA is not capable of determining the diversity of HCPs within a sample, or how the HCP composition changes throughout a manufacturing process, extra methods to complement ELISA need to be used to be able to provide a more in-depth analysis of the HCP population (Bracewell et al, 2015). Due to the drawbacks of antibody-based immunoassays and the problem of identifying problematic HCP species which have the ability to associate with the product, it is crucial to use orthogonal methods that do not depend on immunoreactivity of HCP for HCP detection (Zhu-Simoni et al., 2014). ELISA can be used together with non-specific methods such as 1D- and 2D- polyacrylamide gel electrophoresis (1D/2D-PAGE). Gel electrophoresis is generally used to provide qualitative analysis of proteins within a sample. 2D electrophoresis is more regularly used during process development and characterization, in order to monitor changes in the HCP profile. They have the ability to analyze different HCPs on a single gel. 2D-PAGE gels separate HCPs first on the basis of the isoelectric point followed by separation due to size. The sensitivity of detection can be enhanced by the use of different staining methods which allows visualization of less abundant spots (Tscheliessnig et al., 2013; Hogwood et al., 2014).

Gel-based methods are unable to identify low abundance HCPs, are labour intensive, time consuming and low throughput, also many HCPs could be masked by the product. 2D-PAGE gels can be used in combination with mass spectrometry (MS) to identify the specific HCPs correlating to particular spots on the gel. They have been widely used to examine the HCP composition at different stages of the manufacturing process, either by themselves (Jin et al., 2010) or in combination with mass spectrometry (Tait et al., 2012; Tarrant et al., 2012; Hogwood et al., 2013; Aboulaich et al., 2014; Levy et al., 2014; Valente et al., 2015). Jin et al. (2010) used 2D-PAGE to estimate differences in HCP composition under different upstream and harvest conditions while Hogwood et al. (2013) used 2D-PAGE to investigate HCPs that are present post Protein A from both a mAb producing and null cell lines.

#### **1.4.2.3 Mass spectrometry (MS)**

Tscheliessnig et al. (2013) and Hogwood et al. (2014) presented several orthogonal methods that can be used to complement ELISA as well as 2D-PAGE gels. The method that has been widely used in recent years and appears to be the main complementary tool to ELISA is mass spectrometry. Mass spectrometry has the capability to detect and monitor multiple HCPs, including low abundance proteins, in the same sample, in a short time and in a HT way. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) and surface-enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF MS) are common mass spectrometry methods that either alone or in combination with 2D-PAGE gels have been used to monitor HCPs during process development. In comparison to 2D-PAGE gels that are not high throughput, require large sample volumes and are labour intensive, SELDI-TOF is much faster and requires smaller sample volumes. Both methods give information across distinct mass ranges, SELDI-TOF being able to provide information in the low molecular weight range whilst 2D-PAGE giving information across a higher molecular weight range.

Tait et al. (2012) used SELDI-TOF MS as a comparison method to 2D-PAGE gels, to investigate supernatant HCP profiles at different times throughout the

culture. Valente et al. (2015) used 2D-PAGE gels in combination with MALDI-TOF/TOF MS to identify HCPs in harvest material that showed varied expression across different cell ages. Tarrant et al. (2012) used SELDI-TOF MS to determine the impurity profile within Protein A eluate pools for four different resins. Using this method they identified a number of residual HCPs present after Protein A. Pezzini et al. (2011) used mass spectrometry to determine the HCP profiles in the elution fractions of four different mixed-mode chromatography resins.

Liquid chromatography techniques can be coupled with mass spectrometry (LC-MS/MS) for the identification and quantification of HCPs from CHO cells (Doneanu et al., 2012). Thompson et al. (2014) presented an HCP enrichment method combined with LC-MS/MS which improved the identification and determination of relative abundance of HCPs present in a mAb drug product. After the enrichment step it was possible to identify 19 HCPs, compared to only one before the method was applied. There are also computer-based methods used for HCP identification and risk assessment (bioinformatics portals) (Wang et al., 2009) as well as for analyzing the immunogenicity risk from HCPs in CHO-based protein production (Bailey-Kellogg et al., 2014).

## **1.5 Problematic HCPs**

### **1.5.1 Introduction**

The development of high producing recombinant protein expression systems is causing increased focus to be placed on the product's composition and the downstream processing steps required to remove process-related impurities such as HCPs, lipids, DNA, etc. Even though this is the case, limited knowledge still exists about the range of HCPs that are problematic either by associating with the product throughout the manufacturing process and potentially eliciting an immune response in patients and/or causing product modification. There is also limited understanding regarding the relationship between the protein of interest, bioprocess conditions (both in USP and DSP) and problematic HCPs (Bracewell and Smales, 2013).

After the purification process, low levels of residual HCPs can still be found in the final product (Thompson et al., 2014). The main concern regarding the presence of residual HCP impurities in recombinant protein therapeutics is their potential impact on patient safety, by inducing an unwanted immune response with serious side effects. For this reason, HCP levels need to be reduced to a limit that is considered safe by the regulatory authorities. At present, immunoassays such as ELISA, are almost always used to measure and monitor the total amount of residual HCP, but the composition is generally unknown. The levels of specific, critical HCPs are not independently measured (Zhang et al., 2014). Protein A affinity chromatography is the workforce of the mAb purification process due to its high specificity and capability to remove a large fraction of the total HCP and most other impurities in the supernatant (Butler et al., 2012; Zhang et al., 2014; Gronemeyer et al., 2014). Even after Protein A, problematic HCPs can still be found in the product fraction as well as persisting throughout the whole downstream process until final product (Aboulaich et al., 2014). Recent work has confirmed the need for a better understanding of the presence and potential risk of certain problematic HCPs.

HCPs can be found in the product fraction of bind-and-elute chromatographic steps. This can be due to two main methods. The first method is product-association (Shukla and Hinckley, 2008; Tarrant et al., 2012; Levy et al., 2014; Aboulaich et al., 2014) through strong interactions that specific HCPs have with the mAb, resulting in their binding to the product. These species can then be transported through the DSP process, in combination with the mAb. The second mechanism involves certain HCPs binding to either the resin backbone or the ligand of the chromatographic step, resulting in their elution into the product fraction (Tarrant et al., 2012; Levy et al., 2014; Valente et al., 2015). Out of the two mechanisms, the HCP-mAb interaction seems to be the more common cause by which problematic HCPs are found in the final product. Previous research performed by Tait et al. (2012), Nogal et al. (2012), Sisodiya et al. (2012) and Hogwood et al. (2013) has demonstrated that interactions during Protein A affinity chromatography are generally the main cause of HCPs co-purifying with the product during the purification process. They have also noted that mAbs might interact only with specific subpopulations of the total HCPs (Aboulaich et al., 2014; Levy et al., 2014). Levy et al. (2014) showed that intracellular proteins

(found in cytoplasm, lysosome) represent a large fraction of all proteins interacting with mAbs. This emphasizes how minimizing HCP levels upstream might ease the pressure on downstream process in terms of HCP removal. A subgroup of 118 HCPs have been mentioned in literature as particularly hard to remove during downstream purification due to one of the two described mechanisms (Valente et al., 2015).

Several publications have presented the importance and implementation of wash steps in order to disrupt HCP-mAb interactions during the purification of mAbs using Protein A affinity chromatography (Shukla et al., 2008; Shukla and Hinckley, 2008; Aboulaich et al., 2014). Shukla and Hinckley (2008) compared the efficiency of various intermediate wash buffers in terms of their capability to disrupt HCP-mAb interactions during Protein A and identified a potential wash that can be used as a platform wash condition for Protein A (25 mM Tris, 10% isopropanol, 1 M urea and pH 9.0). Aboulaich et al. (2014) investigated the effects of different wash modifiers on dissociating HCP-mAb interactions. They found that a combination of two wash modifiers (e.g. urea and sodium caprylate; caprylate and arginine) can have the potential to enhance HCP clearance through combined effects, reducing different types of interactions such as hydrogen bond, electrostatic and hydrophobic interactions. Levy et al. (2014) shows that by applying a high-salt wash, product-associated impurities in a protein A affinity chromatography could be removed before product elution.

Process-related impurities must be closely monitored and removed during the downstream process, to ensure that their concentration in the final product does not exceed the general guidelines of less than 100 ppm (ng/mg) for HCPs and 10 ng/dose for DNA (Chon et al., 2011). Important questions arise about what forms/ species of HCPs make up this 100 ng/mg limit and if any of the individual proteins found in this general mass defined as residual HCPs are likely to transform the product through association with it or enzymatic activity. There is reason to believe that this final HCP target might not be acceptable and instead of defining a general limit for all HCPs, there should be a more detailed criteria, based on a better understanding of the HCP population present (Bracewell et al., 2015).

## **1.5.2 Common problematic HCPs**

### **1.5.2.1 Protease**

Unwanted proteolytic activity is well known in mAb manufacturing due to an unknown mixture of proteolytic enzymes likely to be present among the broad range of HCPs present in the harvested cell culture fluid (Sandberg et al., 2006). The issue of proteolytic degradation of recombinant proteins is more common in serum-free cultures, due to the absence of serum proteins which would adsorb the proteases released in the culture medium (Elliott et al., 2003).

There are numerous examples in literature that identify proteases as being problematic HCPs, by either association with the product or due to their impact on the therapeutic protein's structural stability. Sandberg et al. (2006) describes how metalloproteases can destabilize Factor VIII production in CHO cells, Robert et al. (2009) and Dorai et al. (2011) showed how aspartic proteases and serine-threonine proteases are responsible for the fragmentation of fusion recombinant proteins and Gao et al. (2011) reports a high-purity human IgG1 mAb exhibiting fragmentation, that can be associated to residual proteolytic enzyme activity. In more recent studies, Dorai and Ganguly (2014) show how the presence of intracellular and secreted proteases can result in the enzymatic degradation of recombinant proteins, during fermentation and Aboulaich et al. (2014) identified serine protease as a problematic HCP, as it bound to all four mAb investigated and had the potential risk to cause enzymatic degradation of the mAbs. Serine protease was also identified as a purification challenge by a few other publications, which have seen it persist after Protein A chromatography (Doneanu et al., 2012; Hogwood et al., 2013) and mixed-mode chromatography (Pezzini et al., 2011). Wang et al. (2014) outlines the significant impact residual host cell proteases can have on the long-term storage stability of the product. The US FDA raised safety concerns regarding these impurities, by delaying two Phase III clinical trials, as a result of antibodies against residual HCPs being present (Gutierrez et al., 2012).

In order to minimise the impact of degrading enzymes, such as proteases, the addition of inhibitors could in some cases be advantageous (Robert et al., 2009). Dorai et al. (2011) showed how protease cocktail inhibitors as well as inhibitors for a specific protease class (serine-threonine) inhibited the clipping process of a fusion protein. Clipping is a common cause of protein degradation, mostly attributed to the activity of proteases released by cells. The use of inhibitors specific for a particular type of proteases can also help identify the exact variety that is responsible for product degradation. Dorai et al. (2011) also investigated the effect of specific inhibitors for cysteine protease, metallo-protease and aspartic acid protease, each of them with no effect on the clipping process.

### **1.5.2.2 Thioredoxin**

The thioredoxin system or parts of the thioredoxin system (TXN1) have been identified as being the causative factor for the reduction of inter-chain disulphide bonds of monoclonal antibodies produced by CHO cell culture upon scale-up (Kao et al., 2010; Trexler-Schmidt et al., 2010; Koterba et al., 2012). The thioredoxin system is present in the cytoplasm and together with glutathione/glutaredoxin system maintains the cellular redox balance and keeps intracellular protein disulfides reduced. The thioredoxin system is also known as a cell antioxidant (Koharyova and Kolarova, 2008; Kao et al., 2010).

Trexler-Schmidt et al. (2010) firstly identified this problem during harvest operations and determined that the cause was the release of cellular enzymes due to mechanical cell shear. They tested several approaches to try and prevent disulfide reduction and found various levels of chemical inhibitors (EDTA, CuSO<sub>4</sub>, L-cystine), air sparging and low harvest cell culture fluid (HCCF) pH efficiently inhibited the mAb disulfide reduction. Kao et al. (2010) has built on the work of Trexler-Schmidt et al. (2010) and further identified through an *in vitro* experiment, that an active thioredoxin system or other reducing enzymes with thioredoxin-like activity were responsible for the mAb reduction problem in the HCCF. As the main causes for mAb reduction were identified to be the Trx system, hexokinase and G6PD, any inhibitor targeting any of these enzymatic pathways has the potential to prevent

disulfide bond reduction in mAb manufacturing. Koterba et al. (2012) also identified reduction of the mAb's disulfide bonds during large-scale manufacturing of an IgG1 mAb in CHO cells. They established that mammalian thioredoxin 1 (TXN1) is the enzyme responsible for the reduction and showed that by using transduction of a lentivirus expressing TXN1 shRNA, the expression of TXN1 in CHO cells can be reduced and the disulfide bond reduction prevented.

## **1.6 Statistical modelling**

### **1.6.1 Introduction**

High value pharmaceutical products such as vaccines, hormones and monoclonal antibodies are mainly produced using mammalian cell cultures. Culture conditions are highly specialized and by making minor variations to these conditions cells are likely to either become nonviable or have reduced productivity. Mammalian cells have a very complex internal structure where interlinked biochemical processes take place, therefore making an accurate prediction of cell culture behaviour represents a significant challenge (Kontoravdi et al., 2005; Sidoli et al., 2005).

Some of the most important aims of biopharmaceutical industry are the optimisation of cell culture processes in order to maximize antibody production as well as to reduce the time to market. An important task in defining an optimisation strategy is to identify and predict cell culture behaviour (De Alwis et al., 2007). In order to improve antibody production yields, cell culture experimentation is used. This involves trial and error optimisation of cell culture parameters, resulting in a large number of experiments, which can be time-consuming and expensive (Ho et al., 2006). Using mathematical models, initial experimental information can be organized in a coherent manner which helps identify and quantify key relationships between variables, process parameters and product output rates (Bailey, 1998). Parameters that have significant effect on antibody production could be identified by analysing these mathematical models. These can then be singled out for more detailed studies and used to help design experiments. This approach will result in a significant reduction in the required number of experiments with time and cost-saving implications (Ho et al.,

2006). Using mathematical relationships to characterise different parts of mammalian cell culture behaviour and their integration into a predictive model, would contribute to the control of product quality and maximization of antibody production (Sidoli et al., 2004). The aim of mathematical modelling for mammalian cell cultures includes substitution of expensive and time consuming laboratory experiments with *in silico* ones, developing model-based algorithms for controlling product quality (Kontoravdi et al., 2005, 2007), process characterisation and identification of design space (Amit, 2010; Abu-Absi et al., 2010; Rouiller et al., 2012) and the probability of process outputs falling out of specification (Stockdale and Cheng, 2009; Eon-Duval et al., 2012b).

### **1.6.2 Multivariate Data Analysis**

Biotech unit operations are often described by a large number of inputs (operational parameters) and outputs (performance parameters) along with complex interactions between them (Rathore et al., 2014). Given the large amount and complexity of variables in biological systems, it is almost impossible to extract and analyse the information using simple charting, univariate or bivariate methods of analysis. These types of analyses are usually ineffective and likely to result in misleading conclusions (Kourti, 2004; Kirdar et al., 2008). Significant information can be found in the correlations among process parameters and this information is overlooked when parameters are analysed independently (Rathore et al., 2014). This large amount of complex data needs methods of analysis that are able to handle multiple variables simultaneously but also reveal the relationship between them.

Multivariate data analysis (MVDA) can overcome challenges associated with univariate or bivariate analysis such as missing data, variation introduced by deviating factors (noise and experimental error) and multicollinearity (Martin et al., 2002; Kirdar et al., 2008). Multivariate data analysis is the analysis of multiple statistical variables at the same time and it helps understand, visualize and make predictions from the data. MVDA has significant advantages over traditional statistical tools. The powerful data mining abilities allows the analysis of complex data sets to identify essential patterns, while advanced regression methods can be

employed to accurately predict the system's performance. It can predict the effect a change in one variable will have on other variables ([www.camo.com](http://www.camo.com)). Multivariate data analysis in combination with design of experiments can greatly benefit biotech companies by improving process quality and understanding, reducing development timeframes and manufacturing costs and minimising time to market. Multiple linear regression (MLR), principal component analysis (PCA) and partial least squares (PLS) are some of the commonly used projection and regression methods in multivariate data analysis (Rathore, 2007).

PCA is a mathematical analysis that provides a tool for dimensionality reduction leading to better visualisation and quantification of relations between the many variables. It uses an orthogonal transformation to convert a set of observations of potentially correlated variables into new, uncorrelated variables called principal components. The first principal component explains the highest variability in the data, with each following component, which are orthogonal to the previous principal components, accounting for most of the remaining variance (Kirdar et al., 2008; Rathore et al., 2014).

PLS is a regression analysis technique for modelling relations between sets of observed variables. PLS constructs new predictor variables, named latent variables, which are optimal linear combinations of the original explanatory variables. As opposed to PCA which is generally used for the determination of trends, clusters and outliers and process analysis, PLS is used to relate process parameters to process and product quality attributes (Schwartz et al., 2009; Rathore et al., 2011; Abdi & Williams, 2013; Rathore et al., 2014). PCA and PLS are both methods that help compress data by keeping important information and disregarding the noise.

Multiple linear regression (MLR) is a mathematical technique used to model the relationship between multiple independent predictor variables and a single dependent outcome variable (Marill, 2004). Multiple regression has the ability to predict an unknown response variable corresponding to a set of predictor variables as well as understanding the functional relationships between the dependent and independent variables. When multiple regression is used for prediction, the result is

an equation containing partial regression coefficients (McDonald, 2013). MLR provides a linear equation with respect to the predictor variables, but is unable to integrate any non-linear relationships that may occur between the predictors and the response variable. The linearity is only restricted to the model's coefficients therefore the predictor variables can be non-linear. This allows the addition of non-linearly transformed predictor variables (e.g. quadratic terms) in the linear regression modelling. Integrating such variables in the analysis allows for the non-linear behaviour in the data and the interaction between different variables to be taking into account while still having an easy explicable linear model (Hassan et al., 2013).

### **1.6.3 Cell culture models**

Cell culture mathematical models are usually classified as mechanistic (kinetic) and empirical (statistical) models. Empirical models are only expected to accurately describe a set of observation, without taking into account the underlying mechanism. On the other hand, a mechanistic model describes the process, either directly observable or unobservable, under which the data was generated. In predictive microbiology models involve more empirical components (Baranyi and Pin, 2001). Mathematical models can play an essential role in the optimisation and control of bioreactors and fermentation processes.

#### **1.6.3.1 Kinetic models**

Tsuchiya et al. (1966) presented a microbial model classification spectrum that can also be applied to mammalian cell models. In this long established classification system, a mathematical kinetic model can be organized into each of the following categories (Sidoli et al., 2004): unstructured versus structured, unsegregated versus segregated, deterministic versus stochastic. Description of each model type in regards to cell culture modelling is presented in Table 1-5.

#### **Unstructured kinetic models**

Unstructured models are empirical and generally use extracellular data that is typically monitored during a culture, while considering intracellular processes as a

“black box”. Even though this method does not allow in depth study of cellular processes, these kinds of models have been described as being important in supporting optimisation and monitoring strategies (Provost and Bastin, 2004). In addition these models are useful for new cell systems when data is limited (Ho et al., 2006). Quite a few examples of unstructured models used to describe mammalian cell growth (hybridoma cells) have been presented in the literature (Glacken et al., 1989; Frame and Hu, 1991; Xie and Wang, 1994; Zeng, 1996; Jang and Barford, 2000).

**Table 1-5: Classification system of mathematical kinetic cell culture models\***

<b>Classification</b>	<b>Description</b>
Unstructured	Does not take into account the inner structure of the cells
Structured	Incorporates biological knowledge by separating the cells into compartments that are chemically and/or physically distinct
Unsegregated	Cell culture is homogeneous and composed of identical average cells
Segregated	Cell culture is heterogeneous and composed of cells in different stages of development
Deterministic	Cellular processes are not subject to variability
Stochastic	Cellular processes are random

\* Adapted from Sidoli et al. (2004); Kontoravdi et al. (2007a, 2007b)

### **Structured kinetic models**

Structured kinetic models, as opposed to unstructured kinetic models, which do not take into account the inner structure of the cell, incorporate biological knowledge by grouping the biomaterial into distinct compartments (Sidoli, Mataralis and Asprey, 2004). Many structured models have been reported in literature (Tsuchiya et al., 1966; Harder and Roels, 1982; Lee, 2001). As the number of parameters greatly increases with such detailed models, it becomes hard to provide parameter estimation (Flickinger, 2013). A structured, segregated and stochastic model offers the most realistic representation of cell behaviour during cell culture (Sidoli et al., 2004).

### **1.6.3.2 Hybrid models**

Several hybrid models used in chemical and biochemical engineering have been presented in literature throughout the years (Schubert et al., 1994; Roubos et al., 2000; Roubos, 2002; Ho et al., 2006; Kontoravdi et al., 2007). Hybrid models are powerful tools for process monitoring, control and optimisation (Galvanauskas et al., 2004).

### **1.6.3.3 Statistical models**

#### **Kinetic versus statistical models**

Unstructured kinetic models are mechanistic models, which use mathematical and kinetic equations to describe growth rate, substrate consumption and product synthesis. A structured model can be developed from an unstructured kinetic model while splitting the whole process into small parts for a more detailed assessment. In terms of cell culture processes, the metabolism and growth of cells are multistage and complex biochemical processes, thus it is nearly impossible to produce a complete description of the growth and production mechanism. These structured models require the measurement of large numbers of parameters (Baughman and Liu, 1995), therefore are rarely used for the design, control and optimisation of bioprocesses. Statistical (empirical) models are alternatives to conventional model approaches, which are guided by kinetics, material and energy conservation rules (Lee and Gilmore, 2006). One of the disadvantages of statistical models when compared with structured kinetic models is the need for large number of experiments to investigate relationships between factors and responses. A solution to this is the use of design of experiments (DoE) approach, which minimises the number of experiments that need to be performed.

#### **Design of experiments**

Statistical methods require a large number of experiments. In most cases, cost, time and required resources limit the process of generating large amounts of data, therefore the data obtained needs to be rich in information. Using a statistically

designed experiment provides a good solution for acquiring information-rich data from the process of interest (Haaland, 1989, Montgomery, 2004). The design of experiments (DoE) is a computer-enhanced, systematic approach to experimentation that considers all factors simultaneously. The objective of a statistically designed experiment is to obtain effective results in the shortest possible time and with minimum resources (Cornell, 2002, Myers and Montgomery, 2002). Statistical models can be useful when there is no accurate mathematical model equation as well as when there are complex biochemical reactions and interaction between variables. Statistically designed experiments offer the ability to analyse many variables at the same time, with a low number of observations (Lee and Gilmore, 2006).

#### **1.6.4 Predictive modelling**

The introduction of the QbD concept has changed the way of looking at process understanding and control in the pharmaceutical and biopharmaceutical industry. Risk assessments, design of experiments, predictive models, process analytical technology and data analysis are essential tools within QbD. Process modelling is an essential part of the QbD framework (Kourti, 2015). Models are useful in facilitating process understanding and control, determination of design space and process development, making them part of the product lifecycle (Kourti, 2010). Statistical models combined with scale-down HT experimental systems can help improve the ability of gaining a much better design space confidence.

Abu-Absi et al. (2010) developed regression models for specific growth rate and final viability at the seed bioreactor stage as well as normalized titre, high molecular weight (HMW) aggregate species, N-Linked oligosaccharide profile and analytical CEX change variant profiles at the production bioreactor stage. The design space for a monoclonal antibody cell culture process was mapped. The operational parameters at each stage (vial thaw, shake flask inoculum expansion, seed bioreactor and production bioreactor) were prioritized by a risk analysis approach such as failure modes and effects analysis (FMEA) and classified as non-key, key and critical. The model terms with significant effects on titre were seed bioreactor temperature, seed bioreactor final viable cell density (VCD), production bioreactor

temperature shift timing, production bioreactor pH, production bioreactor initial VCD and production bioreactor dissolved oxygen. The resulting regression model for titre had a reasonable prediction capability (predicted  $R^2$  of 0.61). Using predictive models, process conditions were optimised in order to achieve high product titre while controlling glycosylation.

Rouiller et al. (2012) generated models for product titre, bioactivity and the levels of various product variants (glycoforms, oxidized forms), product-related impurities (HMW species, clipped forms) and process-related impurities (HCP, DNA). The production bioreactor step of an Fc-Fusion protein manufacturing cell process was characterized following quality by design (QbD) principles. A risk assessment exercise was employed to identify potential critical and key process parameters with a possible impact on product quality and process performance. Analysis of variance (ANOVA) was used to generate models for titre and each of the critical quality attribute identified using Design Expert software. The parameters influencing product titre were pH, dissolved oxygen (DO), culture duration and seeding density, while those influencing HCP were pH and to a lesser extent dissolved oxygen and culture duration. The resulting models were used to define the cell culture design space. The design space established for cell culture process resulted in the establishment of operational ranges for pH, DO and cell culture duration in order to provide consistent delivery of a cell culture harvest that meets the DSP requirements for impurities and the drug substance.

## **1.7 Bioprocess economics modelling**

### **1.7.1 Introduction**

The success of new biopharmaceutical candidates relies more and more on economic issues. Pressures for cost and time reduction of drug development lead to links being made between a company's manufacturing strategy and business strategy (Farid et al., 2005b). In order to minimize cost of development, reduce time-to-market and quantify risks for maintaining economic returns, fast and effective tools are necessary. Computer-aided design tools are able to help achieve these objectives and

guarantee rapid delivery of drugs to patients. The use of software tools to help make business decisions is critical. The application of process-cost modelling enables rapid comparison of different manufacturing and development options (Lim et al., 2010). Simulations can be used for cost analysis, resource utilization and mass balance assessment and are an important part in the analysis and selection of process options and characterization of unit operations, contributing to better decision-making in terms of business and process needs (Lim et al., 2005).

### **1.7.2 Factors impacting COG/g**

In order to lower COG/g, efforts are made to try and decrease batch costs or to increase the overall process productivity by an increase in step yields. Increases in titres are expected to have a significant impact on lowering COG/g, assuming that DSP costs do not countermand the improvements in USP. As titres increase, the main costs of the manufacturing process are shifted to the DSP process as high protein loads are needed to be purified by the chromatography steps (Butler and Meneses-Acosta, 2012). This would result in adopting higher number of cycles or purchasing larger columns in order to cope, leading to an increase in costs per batch. The overall COG/g could still be reduced if the increase in the overall productivity compensates the higher DSP costs. As mAb titres will increase further in future years, the purification process will become a significant contributor to the overall COG/g and will drive the need for optimisation and cost savings (Farid, 2009).

The number of downstream steps as well as the yield of each step has an impact on the overall DSP yield. Increases in step yields combined with a reduction in DSP steps have led to a 40-75 % increase in overall yield, contributing to lower cost of goods (Werner et al., 2004; Li et al., 2005). Sommerfeld and Strube (2005) highlighted the importance of individual step yield increase on the overall COG/g. They showed that increasing the average step yield in a seven-step process from 85 to 95 % led to an increase in the overall yield from 30 to 70 %, corresponding to a 40 % lower downstream COG/g. In mAb manufacturing, the overall yield could be improved through the elimination of buffer exchange steps, by developing chromatography steps able to process material eluted from the previous step, with no

further treatment. Higher protein loads resulted from high titre cultures has an impact on the capacity of chromatography columns, especially when the limit for column diameters is 2 m. More cycles required to cope with the increased load will result in longer processing times, reducing the facility throughput and influencing COG/g. Improvements in resin binding capacity would reduce resin volumes and amount of buffer used. This would reduce consumables costs, a main part of the COG/g at higher demands and titres (Sommerfeld and Strube, 2005; Farid, 2009).

### 1.7.3 Classification of process economic models

Standard process economic models deal with problems such as predicting COGs, capital investment and cash flow analysis as well as risk assessment and project management. In order to determine COGs for a process, process models need to be combined with cost models (Farid et al., 2000). Certain parameters from process models such as resource usage and overall output are essential to COG models (Farid et al., 2005a, Mustafa et al., 2006). When selecting an appropriate software platform for these models, factors such as the desired outputs, the tool's requirements specification and the level of detail need to be considered. Based on which decisions are made, the model can be static or dynamic and deterministic or stochastic (Table 1-5).

**Table 1-6: Classification system of process economic models**

<b>Classification</b>	<b>Description</b>
Static	Unable to account for time dependent variations
Dynamic	Allow the evaluation of simultaneous processes over time Enable a more realistic estimation of cost
Deterministic	Process outputs do not account for risk
Stochastic	Incorporate uncertainty within outputs Give a more realistic overview of the process

### **1.7.3.1 Static versus dynamic models**

Static models are the most commonly used models, spreadsheet-based, simple and easy to build. Static process models can be combined with cash flow models, to work out how profitable investments are as well as to COGs models in order to obtain the cost breakdown and evaluate how sensitive COG is to different process parameters. These can be best applied to estimate costs in early stages of projects. Puich and Paz (2004) showed that static models are not suited for situations where delays occur due to resource constraints.

Dynamic models are more complicated to build and use compared to static models, but these provide a more realistic tool for outputs estimation such as cost and throughput. They are time-dependent therefore they can evaluate how operations will change over time. Dynamic models are able to analyse simultaneous events, logistics and delays taking place during manufacturing process due to resource limitations (Puich and Paz, 2004, Rathore et al., 2004, Farid, 2007a). Dynamic process models are best implemented using discrete -event software packages while COG and cash flow are better observed in spreadsheet-based software (Mustafa et al., 2004, Lim et al., 2005, Lim et al., 2006).

### **1.7.3.2 Deterministic versus stochastic models**

Deterministic models are conventional process models, which assume the process is certain if the inputs are fixed. The process outputs do not account for risk that might occur. This is not realistic as a manufacturing process is always facing technical and market-related uncertainties such as batch failures, cell culture titres, purification yields and processing time (Biwer et al., 2005, Lim et al., 2006).

Stochastic (or probabilistic) models are models in which some of the model's components are taken from a probability distribution. The outputs of these models are able to capture uncertainties giving a more realistic overview of the process studied (Wang, 2010). An overview of the deterministic and stochastic bioprocess economics models presented in the next section are shown in Fig. 1-5.

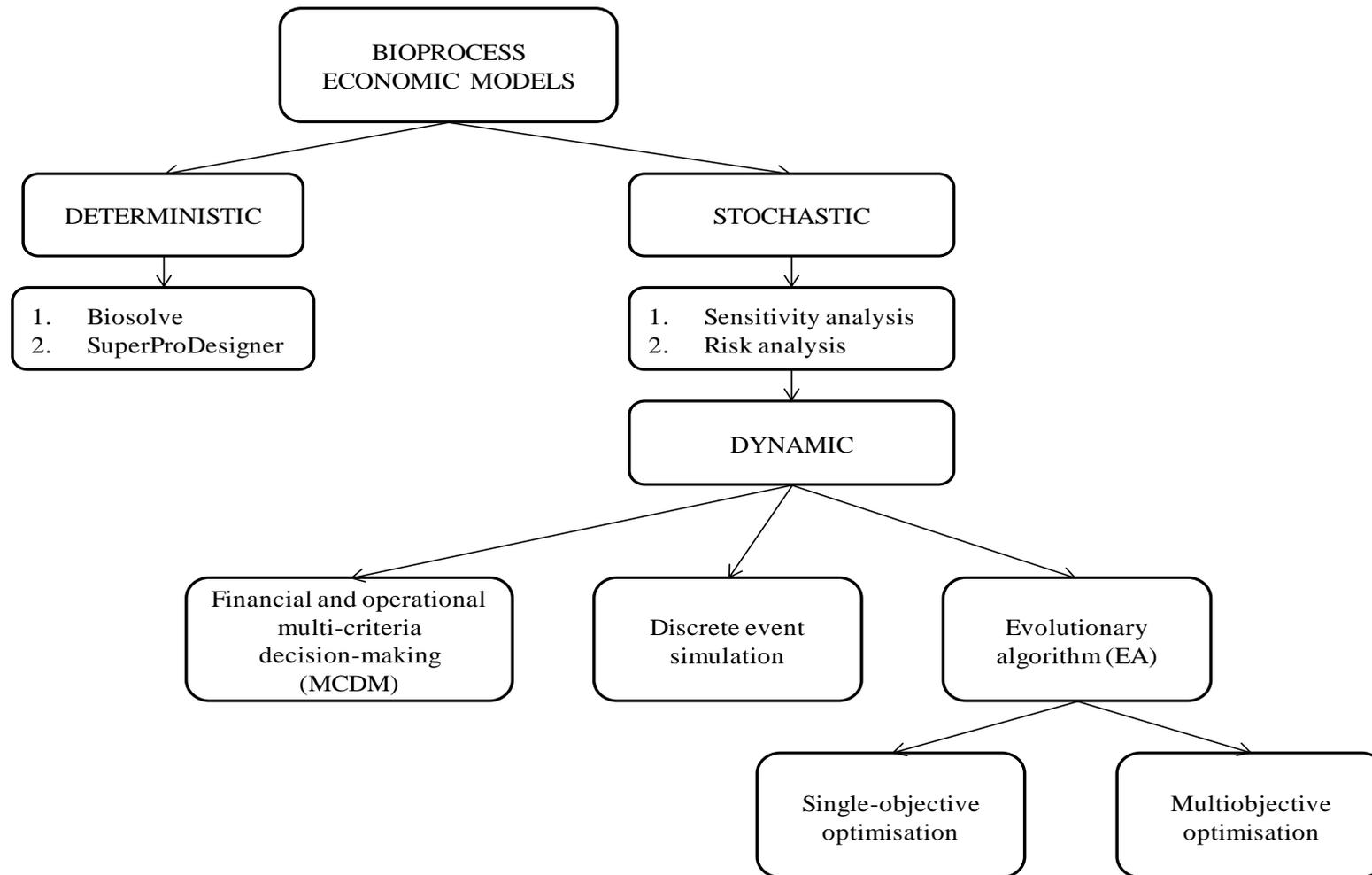


Figure 1-5: Overview of deterministic and stochastic bioprocess economic models presented in this section

## **1.7.4 Deterministic process economics**

### **1.7.4.1 Biosolve**

Biosolve is a Microsoft Excel-based process-cost modelling tool that can be used for the optimisation and development of cost-effective bioprocesses. It can be applied in early stages of process development in order to recognize the impact of scale, as well as financial impact and to help make better business decisions. Biosolve can also be used to examine the most recent technologies and analyse their impact on specific processes ([www.biopharmservices.com](http://www.biopharmservices.com)). The software is easy to use particularly by users that are already familiar with Microsoft Excel and it has a flexible configuration. Within the cost model a sequence of unit operations is defined. Process parameters such as flux rates and binding capacity can be specified as well as personnel requirements and operating time. The effects of varying different parameters on the operating costs can be analysed through two analysis tools: Sensitivity and Scenario. A cost database is included within the model to maintain costs regarding labour, consumables, etc. The main outputs of the software include capital investment, materials cost, all of which play a part in the COG calculation (Lim et al., 2010).

Sinclair (2010) used the Biosolve software to assess the impact of different geographic locations on the COGs of a standard monoclonal antibody process. A case study is used to estimate cost contributions of different single-use systems on the manufacturing costs, making the comparison with stainless steel systems by taking into account various geographical regions. It was observed that the capital required to build a stainless steel facility was lowest in China, followed by Singapore, Japan and the UK. This is due to the lower labour fees paid in China. A 29 % reduction is observed when comparing the UK with China. When changing from stainless steel to disposables, similar savings are occurring in each region. It is observed that the amount of savings is lowered as manufacturing costs are reduced.

#### **1.7.4.2 SuperPro Designer**

SuperPro Designer is a commercial batch simulation package used to model bioprocesses. It involves all the process and subprocess steps as well as the cost associated with them. The generated COG models allow accurate estimation of the final cost of the product produced at manufacturing scale – from vial thawing to final product (Costioli et al., 2010). Some of its advantages include the ability to estimate profitability and cost, to quickly scale-up or scale-down equipment sizes for various annual outputs and the ability to generate graphical representation of flowsheets. This tool is able to simultaneously perform equipment sizing and costing, economic evaluation and material and energy balances (Shanklin et al., 2001). Some of the limitations associated with static models such as SuperPro Designer include the inability to account for constraints such as resource and utility as well as handling large datasets (Stonier et al., 2012). Another downside is the fact that probability distributions cannot be used to incorporate and illustrate uncertainties associated with parameters (Mustafa et al., 2006).

SuperPro Designer has been used to estimate the production cost, capital investment and key profitability indicators of monoclonal antibodies production processes using mammalian cell culture. Oh et al. (2004) used this tool to investigate the impact caused by increasing the staggered number of fermenters to double the annual production rate. A comparison between the base case and the optimized case showed that by increasing capacity, the total capital investment increases by 12% and the annual production cost increases by 88% while the ROI increases by 78% and the unit production cost decreases by 7%. Harrison et al. (2003) demonstrated how sensitive the production cost is to the annual production rate and showed the exponential decrease of the production cost as the production rate is 10 times higher. It has also been shown that the annual production rate influences the upstream to downstream costs ratio, therefore when the production rate increases, the costs move towards the downstream processing.

### **1.7.5 Stochastic process economics**

It is difficult to manage the manufacturing of biopharmaceuticals in order to maximize throughput and minimize the COGs due to technical, clinical and commercial uncertainties. There is an increased interest in the ability to model uncertainty in manufacturing operations. The most common uncertainties influencing the manufacture of biopharmaceuticals are technical and market-related. Some of the technical uncertainties include downstream processing yield, duration of manufacturing tasks, product titre during cell culture and the possibility of contamination. Market uncertainties comprise of dosage levels, costs of resources and market demand (Lim et al., 2005; Stonier et al., 2012). There are different ways of considering uncertainty such as sensitivity analysis and risk analysis including risk adjusted values and Monte Carlo simulations. These methods are described below.

#### **1.7.5.1 Sensitivity analysis**

Carrying out a sensitivity analysis of the principal variables is the easiest method of determining and quantifying uncertainties related with a particular project. Sensitivity analysis gives a systematic way to examine the effect of changing parameters by determining the impact of  $\pm x$  % changes in each variable on the output measures therefore establishing the stability of the base case (Lim et al., 2005; Thabane et al., 2013).

#### **1.7.5.2 Risk analysis**

Risk analysis is an important part of Process Analytical Technology (PAT) used in the biopharmaceutical industry. PAT is an essential tool for the implementation of QbD, which can be used to monitor and control the manufacturing process (Riley and Li, 2011). Methods of incorporating risk necessary for more complicated problems require the assessment of probability functions for key uncertain factors. This can be done through ‘risk adjustment’ and ‘Monte Carlo simulation’ (Farid et al., 2005; Farid, 2007a). In risk adjustment, each input is weighted by how likely it is to occur. The output measures are then risk-adjusted values that illustrate expected average values, which consider all possible outcomes.

## **Monte Carlo Simulation**

Monte Carlo simulation is a type of stochastic modelling that can be used to determine the impact of project uncertainties on the outputs. It uses the input probability distributions to find out the probability distributions of the outputs. Monte Carlo technique imitates the randomness inherent in manufacturing by generating random outcomes for probabilistic factors when applied to a static or dynamic model. Repeating this simulation process a large number of times leads to a range of possible output values that help determine the system's performance. Monte Carlo simulation has been used in various bioprocessing economic studies in order to examine the impact of manufacturing uncertainties such as batch failure, product titre and yield on cost (Pollock et al., 2013; Simaria et al., 2012; Allmendinger et al., 2014b). Gold Sim (GoldSim Technology Group LLC, Washington, USA) and @RISK (Palisade Corporation, Newfield, NY, USA) are some of the commercial packages used for Monte Carlo simulations.

### **1.7.6 Dynamic and stochastic studies**

Upstream process decisions are generally based on either perfusion or fed-batch cultures. Lim et al. (2006) used a risk-based discrete-event tool to assess the economic feasibility of fed-batch and perfusion cultures via a case study based on commercial production of antibodies at the 50 kg scale. The lower productivity and higher start-up costs of fed-batch processes were compared to the higher productivity and higher operational risks of perfusion processes. The deterministic analysis showed that the annual COG/g were very similar for both options under the stated assumptions, with 3% reduction in the perfusion option compared to the fed-batch. The deterministic analysis (no risk) recommended the perfusion option as the more economically feasible option due to the higher projected net present value (NPV) and lower initial investment. The stochastic analysis (incorporating risk by accounting for fluctuations in product titre, DSP yield and the feasibility of contamination and equipment failure) showed that the perfusion option had a lower reward/risk ratio and failed to meet the expected output. This would make the perfusion option infeasible when accounting for uncertainties and risks. The studies presented

highlighted the importance of incorporating risk and uncertainty when making manufacturing and economical decisions and the limitations of relying on deterministic analysis alone.

#### **1.7.6.1 Financial and operational multi-criteria decision-making (MCDM)**

Multiattribute decision making (MADM) models involve making decisions in the presence of various, usually conflicting criteria (Rao, 2007). The models have the ability to include qualitative as well as nonfinancial aspects of performance in the evaluation of different decisions (Farid et al., 2005b). The use of multiple criteria can be supported through multi-criteria decision-making (MCDM), one of the most common sections of decision-making (Triantaphyllou, 2000). A variety of methods are available in MCDM that have been used to process non-financial and financial data. For example, Platts et al. (2002) used the additive weighting technique to analyse the decision to invest in internal manufacturing capabilities or to outsource, Farid et al. (2005b) used the same method to investigate the decision to either build a pilot plant with only stainless steel equipment, a pilot plant based on fully disposable components or a hybrid pilot plant with stainless steel fermenters but with fully disposable components in the downstream production areas. Furthermore, Steuer and Na (2003) published a review of 265 publications that concentrate on using MCDM to assist decision-making in financial situations.

George et al. (2007) presents the development of a decision-support framework for decision-making scenarios that uses multi-criteria decision-making (MCDM). Its functionality is demonstrated through a case study based on a biopharmaceutical company confronted with several options for purchasing commercial manufacturing capacity. A stochastic analysis of options was carried out. The framework was implemented in Microsoft Excel and is composed of four features: a biomanufacturing process model, a profit and loss model, a MCDM technique (additive weighting technique) and a number of criteria used to differentiate between the different options. The model's input variables are the expected fermentation titre, the anticipated success rate of each batch, the annual demand and the overall product yield, whereas the model's outputs are COG/g and

the fixed capital investment (FCI) for constructing the plant. The deterministic analysis showed that the *Build*, *Partner* and *Partner/Build* options scored higher than the average additive weighting technique score. The *Build* option was the preferred option after this analysis as it had the highest score, followed by the *Partner* option and the *Partner/Build* option. This option results in the highest NPV value and the highest total value of assets. The sensitivity analysis showed that the most significant factor influencing the deterministic results was the company's market capture. Monte Carlo simulation has demonstrated that the most profitable option and the one that has a greater potential to generate profit was the *CMO/Build* option. A stochastic analysis proved that the ranking positions remain the same as for the deterministic analysis with *Build*, *Partner* and *Partner/Build* being the top three. The application of the model to a case study highlighted the limitations of using a single criterion when making strategic manufacturing decisions as other important criteria might have been omitted. For the best option to be identified, the use of multiple criteria analysis under uncertainty is critical.

#### **1.7.6.2 Discrete-event simulation**

Stonier et al. (2013) presents the implementation of a decision-support tool designed to resemble process variations using advanced multivariate statistical techniques, to help discover the reasons of short-term facility fit problems. The large datasets generated from biopharmaceutical industrial batch processes are evaluated using principal component analysis together with clustering algorithms. This study expands on an already developed database-driven simulation platform that includes process economics, equipment sizing and mass balancing of purification sequences in antibody manufacturing processes (Stonier et al., 2012). This study presents the expansion of the tool to be able to mimic the stochastic aspects of industrial batch processes achieved by creating the ability to perform Monte Carlo simulations and identifying what is the best way to incorporate the stochastic results into advanced multivariate statistical analysis techniques. A typical monoclonal antibody manufacturing process was simulated in a 10.000 L facility. Generation of the base case data considering a titre of  $2 \pm 0.2$  g/L shows that the predicted throughput is below the required one. This suggests a potential facility fit problem. The simulation

tool registered an error event when the tanks volume is exceeded. This arose in the AEX chromatography and virus inactivation pool volumes. As the largest volume that these tanks can hold is 2500 L, the extra volume was sent to waste resulting in product losses. These losses have a big influence on the overall throughput when a high number of batches are performed. This facility fit problem could be resolved by modifying the facility to redirect the product into auxiliary tanks. After the process modification, the variation in mass loss is removed.

### **1.7.6.3 Evolutionary (genetic) algorithms (EA)**

#### **Single-objective optimisation**

Optimisation techniques are used when the number of scenarios to be explored is too large for the scenarios to be individually evaluated. Simaria et al. (2012) introduces a meta-heuristic optimisation approach using genetic algorithms to focus on the difficulty of designing facilities with several possible permutations. This single-objective optimisation technique is able to address various decisions at the same time in order to decide how best to design these facilities and minimise COG/g. This paper focuses on the design of flexible and cost-effective facilities while considering various purification sequences per product. The design of these facilities is regarded as an optimisation problem in which choices made at levels such as product, facility and unit operation represent the decision variables. The problem consists in determining the equipment sizing for each operation, the optimal upstream to downstream trains ratio as well as the sequence of purification steps that needs to be used for each product, while minimising COG and maintaining purity and demand for various products having different yields, demands and impurity levels. The algorithm is connected to a detailed process economics model to examine the numerous operational and financial outputs of each option. The applicability of this algorithm is demonstrated through an industrially relevant case study. The case study examines the design of optimal purification sequences and chromatography column sizing strategies for a manufacturing facility producing three mAbs in different stages of development, with different demands and titres. The tool allowed the

selection of the most cost-effective purification sequences and sizing for each product in the facility.

Allmendinger et al. (2014a) builds on the work of Simaria et al. (2012) and examines the application of evolutionary algorithms for the identification of chromatography column sizing strategies for the sequence of purification steps used in the purification of mAbs. The closed-loop optimisation problem was defined as single-objective (minimise COG/g), subject to various constraints and uncertain parameters. Monte Carlo (MC) simulations based on probability distributions were adopted in order to account for the impact of uncertainty (fluctuations in titre) on COG/g. To demonstrate the framework's ability to identify cost-effective chromatography equipment sizing strategies, an industrially-relevant case study looking at a single-product mAb facility was used. Chromatography column sizing strategies that resulted in savings up to 20 % in COG/g compared to the common approach used by industry (base case) were identified by the algorithm. Using stochastic EAs allowed the determination of more robust solutions, able to handle titre fluctuations.

### **Multiobjective optimisation**

Allmendinger et al. (2014b) presented a framework which linked an evolutionary multiobjective optimisation algorithm (EMOA) to a process economics model. The aim was to identify sequences of chromatography purification steps and column sizing strategies that are subject to multiple objectives including COG/g, robustness in COG/g and the ability to remove impurities. An industrially relevant case study with different demands, USP: DSP train ratios and HCP levels was used to show the framework's ability to identify purification processes that satisfy the objectives and are robust to uncertainty. The uncertainty was modelled using probability distributions during Monte Carlo trials in factors such as product titre, eluate volumes, dynamic binding capacities, step yields, HCP log reduction and initial HCP. The aim was to understand the influence of uncertainty on the DSP design with regards to chromatography sequence and column sizing. Variations in step yields and product titre have the most significant impact on COG/g while variations in initial

HCP levels and HCP logs impacted on HCP level post purification. The framework was able to identify purification processes that offered savings up to 10% compared to the industrial platform.

Making good decisions early in the development cycle of biopharmaceuticals is critical to the success of this sector. The use of computer-assisted models which are able to integrate bioprocess economics with stochastic behaviour, multiple conflicting objectives, manufacturing logistics and numerous constraints have been presented. Decision-support tools provide a very important resource to use in assessing alternative strategies to handle future challenges (Farid, 2013).

## **1.8 Aims and Organisation of thesis**

The previous sections of this chapter outlined the current-state and future directions of biopharmaceutical drug development focusing on monoclonal antibodies. Overviews of the impurities encountered in biopharmaceutical manufacturing (with an emphasis on host cell proteins) including the challenges they pose were presented. In addition, statistical and economics modelling techniques currently used in evaluating biopharmaceutical manufacturing process were also highlighted. In the literature review it was shown that the antibody sector has made significant progress in increasing cell culture titres. However there is a limited understanding of the consequences of mAb titre increases on impurity levels and the subsequent downstream processing performance. Therefore, it is essential to have systematic methods to explore such interactions. Also, the impact of cell culture conditions on the levels of certain problematic HCPs (e.g. protease) as well as the relationship between HCP levels and protease activity has not been previously investigated.

The aim of this thesis was to develop a systematic framework based on QbD principles, combining state-of-the-art, DoE driven, high throughput cell culture experiments (ambr system) with statistical cause-and-effect predictive correlations and process economic models. This will facilitate the identification of cell culture strategies that balance the needs of upstream and downstream manufacturability, robustness to process fluctuations and cost-effectiveness, early in the development cycle. An additional aim was to provide a better understanding of the relationship

between the protein of interest (mAb), cell culture conditions and the levels of HCPs and certain problematic HCPs (proteases) present in the harvest material. In order to achieve these aims, certain objectives were formulated and these form the basis of the subsequent chapters.

Chapter 2 provides an overview of the materials, equipment and analytical techniques used to perform small-scale cell culture experiments and protease studies. Statistical methods used to derive and evaluate predictive correlations developed based on high throughput ambr are presented. Software used to facilitate the statistical and economics analysis of cell culture strategies was also shown.

Chapter 3 presents a QbD approach to cell culture process development (high throughput ambr experimentation and Gyrolab analytics linked with DoE) to characterise cell culture performance associated with different generation numbers and explore the consequences of titre increase on HCP levels at harvest.

Chapter 4 investigates the impact of cell culture parameters (temperature, media osmolality and seeding density) on protease activity at harvest as well as examines the relationship between HCP levels, mAb concentration and protease activity resulting from unclarified harvest samples. A commercially available protease assay was optimised in order to make it suitable for the analysis of unclarified cell culture harvest.

Chapter 5 presents the use of multivariate analysis techniques (multiple linear regression) to characterise the high throughput cell culture data generate in Chapter 3 using the ambr system and derive predictive cause-and-effect correlations. The statistical equations are able to predict cell culture outputs (mAb titre and HCP levels) based on cell culture inputs (temperature, seeding density, media osmolality, pH and timing of feed initiation).

Chapter 6 explores the integration of the predictive modelling equations derived in Chapter 5 with a prototype bioprocess economics and optimisation tool in order to identify the most cost-effective cell culture strategies as well as the impact of uncertainty in cell culture parameters on outputs (product output (kg) and HCP<sub>final</sub> (ng/mg)) and the likelihood of these falling out of specification.

Chapter 7 presents a summary of the main conclusions and discusses possible directions for future work.

## **Chapter 2**

### **2 Materials and methods**

#### **2.1 Equipment used**

##### **2.1.1 ambr system**

The ambr 24 and ambr 48 systems were used, consisting of 24/48 single-use bioreactors split into two/four cultures stations (CS1, CS2, CS3, and CS4) each containing 12 micro bioreactors. The ambr 24 was used for generating AMBR 1, AMBR 2 and AMBR 3 experimental data while ambr 48 was used to generate AMBR 4 experimental data. Each culture station has independent stirring and temperature control. Each micro bioreactor is equipped with a miniature marine impeller, a sparger for gas supply, as well as integrated optical sensors for pH and dissolved oxygen (DO) providing individual closed loop control of these parameters. In order to maintain aseptic operations during the culture, the system is placed inside a bio-safety cabinet (Lewis et al., 2010; Hsu et al., 2012; Moses et al., 2012).

#### **2.2 Cell culture**

##### **2.2.1 Cell line**

The cell line used was a proprietary MedImmune CHO high producing cell line, expressing an IgG1 antibody. The cell line was cultured in protein-free CHO media supplemented with a two-part proprietary nutrient feed. Cells from different cell stocks (presenting different generation numbers) have been used for each of the AMBR experiments: 33, 44, 54 and 59 for AMBR 1, 2, 3 and 4 respectively. The cells with a generation number of 33 within AMBR 1 are closer to a mid-generation number as used by MedImmune for this particular cell line, whereas cells with a generation number of 54 within AMBR 3 are closer to a late generation number. For the purpose of this thesis, experiments with a generation number of 33, 44 and 54 will be referred to as low, mid and high generation.

### 2.2.2 Fed-batch protocols

Three levels of osmolality were used throughout the experiments: base level, intermediate and high osmolality media. The base osmolality refers to the osmolality of the proprietary MedImmune media, used for routine cell culture. The intermediate and high osmolality conditions were prepared by the addition of different amounts of NaCl stock solution to the base medium, while the concentration of all other components was kept constant. One day prior to inoculation 8 mL of media of different osmolalities was added to each ambr vessel, according to the DoE design. Prior to ambr inoculation, different shake flasks were set up with double the seeding density intended within the ambr DoE design, in the corresponding media osmolality. An 8 mL volume was then added from each shake flask to the corresponding ambr vessel to make up a final volume post inoculation of 16 mL with the correct seeding and media osmolality.

Feeding of the cultures started on day two with five subsequent additions that involved the addition of a two-part feed (average 400  $\mu\text{L}$  Part A feed and 24  $\mu\text{L}$  Part B feed). For all ambr cultures, glucose concentration was monitored throughout and fed up to 8  $\text{g L}^{-1}$  when the concentration fell below 5  $\text{g L}^{-1}$ . A sample of 100  $\mu\text{L}$  was taken every day from each ambr microbioreactor for the measurement of glucose and lactate concentration. Starting from the media addition day, 20  $\mu\text{L}$  of antifoam was added every other day to each AMBR culture. Cell counts measurements were performed every day using ViCell. A sample of 600  $\mu\text{L}$  was taken on day 0 followed by a  $\frac{1}{4}$  dilution with PBS for the subsequent days until harvest (150  $\mu\text{L}$ ). Titre and HCP samples (500 and 400  $\mu\text{L}$ , respectively) were taken in the last five days of culture (day 11-15). Offline pH measurements (400  $\mu\text{L}$ ) were performed every other day starting with day 1. For all cultures, pH was controlled within  $\pm 0.1$  of the set value for each culture using  $\text{CO}_2$  in the inlet gas and the DO was maintained at 50 % of air saturation. According to the vendor guidance, a working volume of between 10-15 mL in ambr 24 is sufficient in order to maintain optimum gas exchange and mixing time. Taking into account all the sample and addition volumes, the working volume within all of the ambr experiments did not drop below 10 mL until day 14. From day 13-15, the volume dropped from 10 mL to 8.1 mL. As the working volume

was maintained over 10 mL until day 14, it is not believed that the sampling and additions had any impact on the growth of the cultures (peak viable density usually achieved in day 8-10). Differential dilution across different conditions could only come from the small differences in glucose addition based on the algorithm presented above and base additions during the culture. There were very small differences in final volumes between different conditions which suggests that the differences were negligible.

## **2.3 DoE experiments**

### **2.3.1 AMBR 1**

A DoE design with two parameters (seeding density and media osmolality) varying on three levels ( $3^2$  full factorial) was set up for each of the two culture stations within AMBR 1, using low generation cells (33). The micro bioreactors were seeded at a density of 0.49, 0.8 and  $1.14 \times 10^6$  cells mL<sup>-1</sup>, respectively and osmolalities of 314, 353 and 394 mOsm kg<sup>-1</sup> were used. A standard culture temperature of 36.5 °C was maintained for all cultures within CS1, while in CS2, a temperature shift to 33 °C was carried out after day 4 of culture. Certain cultures within AMBR 1 and AMBR 3 were identified as outliers using PCA analysis. These cultures also showed an abnormal lactate profile compared to the rest of the cultures (data not shown) and were removed from subsequent analysis. The experimental design for AMBR 1 as well as the exact number of outliers removed from each culture station within each ambr experiment is summarised in Chapter 3, Table 3-1.

### **2.3.2 AMBR 2**

For each culture station within AMBR 2, a  $3^2$  full factorial DoE design was carried out, using mid generation cells (44). For CS1, the seeding density and pH were varied on three levels (0.42, 0.69 and  $1.06 \times 10^6$  cells mL<sup>-1</sup>; 6.6, 6.8 and 7) while the addition of feeds started on day 2 of culture for all 12 micro bioreactors. For CS2, all micro bioreactors were seeded at a density of  $0.9 \times 10^6$  cells mL<sup>-1</sup>, while pH and timing of feed initiation were varied on three levels (6.6, 6.8 and 7; feed start day 1, 2 and 3). A temperature of 36.5 °C and an osmolality of 313 mOsm kg<sup>-1</sup> was

maintained for all 24 experiments. The experimental design for AMBR 2 is summarised in Chapter 3, Table 3-1.

### **2.3.3 AMBR 3**

A DoE design with two parameters (seeding density and media osmolality) varied on three levels ( $3^2$  full factorial) was set up for each of the two culture stations within AMBR 3 using high generation cells ( $54$ ). The micro bioreactors were seeded at a density of  $0.58$ ,  $1.19$  and  $1.8 \times 10^6$  cells  $\text{mL}^{-1}$ , respectively and osmolality levels of  $310$ ,  $356$  and  $389$   $\text{mOsm kg}^{-1}$  were used. A standard culture temperature of  $36.5$  °C was maintained for all cultures within CS1, while in CS2, a temperature shift to  $33$  °C was carried out after day 4 of culture. The experimental design for AMBR 3 is summarised in Chapter 3, Table 3-1.

### **2.3.4 AMBR 4**

An ambr 48 was used to set up a face centred full factorial DoE design with centre and axial points at two temperatures. The samples from these experiments were used to carry out assay development work as well as generating results for protease analysis. Within the DoE design, three numerical factors were varied on five levels, seeding density ( $0.53$ ,  $1.1$ ,  $1.6$ ,  $2.1$  and  $3.57 \times 10^6$  cells  $\text{mL}^{-1}$ ), media osmolality ( $311$ ,  $323$ ,  $335$ ,  $350$ ,  $365$   $\text{mOsm kg}^{-1}$ ) and pH ( $6.6$ ,  $6.8$ ,  $7$ ,  $7.2$ ,  $7.4$ ). A standard culture temperature of  $36.5$  °C was maintained for all cultures within CS1 and CS2, while for CS3 and CS4, a temperature shift to  $33$  °C was carried out after day 4 of culture.

## **2.4 Analytical techniques**

### **2.4.1 CCH, $q_{\text{Mab}}$ and specific HCP productivity ( $q_{\text{HCP}}$ )**

Cell population growth was assessed in terms of the cumulative cell hours ( $10^6$  cells  $\text{mL}^{-1}$  h). CCH was calculated by summing the areas under the viable cell growth curve. Each area (corresponding to the area between successive cell counts) was calculated as follows:

$$\text{CH (viable cells)} = \left( \frac{\text{VCD}_0 + \text{VCD}_1}{2} \right) \times (h_1 - h_0) \quad (1)$$

where  $\text{VCD}_0, \text{VCD}_1$  represent the viable cell density (cells  $\text{mL}^{-1}$ ) in day 0 and day 1 whereas  $h_0, h_1$  represent the elapsed time (in hours) between two cell concentration readings. CCH was then calculated by summing the areas underneath the entire growth curve from inoculation to harvest:

$$\text{CCH}_0 - n \text{ (viable cells)} = \text{CCH}_0 + \text{CCH}_1 + \dots + \text{CCH}_n \quad (2)$$

The  $q_{\text{Mab}}$  at harvest was calculated as follows:

$$q_{\text{Mab}} \text{ harvest} = \left( \frac{\text{Harvest Antibody Titre}}{\text{Harvest CCH (viable cells)}} \right) \times 24 \quad (3)$$

The  $q_{\text{HCP}}$  at harvest was calculated as follows:

$$q_{\text{HCP}} \text{ harvest} = \left( \frac{\text{Harvest Host Cell Protein}}{\text{Harvest CCH (viable cells)}} \right) \times 24 \quad (4)$$

#### 2.4.2 Cell generation number

$$\text{Gen no } 1 = \frac{\ln \left( \frac{\text{VCD}_{(\text{passage})}}{\text{VCD}_{(\text{seeding})}} \right)}{\ln (2)} \quad (5)$$

where  $\text{VCD}_{(\text{passage})}$  represents the viable cell density (cells  $\text{mL}^{-1}$ ) of the day of passage (each 3 days) and  $\text{VCD}_{(\text{seeding})}$  represent the seeding density =  $0.3 \times 10^6$  cells  $\text{mL}^{-1}$ . The final generation number was calculated by adding the increase in generation number with each passage to the generation number of the previous passage.

$$\text{Gen no } 1 - n = \text{Gen no } 1 + \text{Gen no } 2 + \dots + \text{Gen no } n \quad (6)$$

### **2.4.3 Viable cell concentration and Viability**

Viable cell concentration and percent viability were determined by the trypan blue exclusion method using Vi-Cell™ XR Cell viability analyzer (Beckman Coulter, High Wycombe, UK).

### **2.4.4 Protein A HPLC Analysis**

Monoclonal antibody concentrations were determined by protein A High Pressure Liquid Chromatography (HPLC) using an Agilent 1200 Series HPLC system (Agilent Technologies, South Queensferry, UK). Various sample volumes were loaded onto a Poros® Protein A 20 µm Column (Applied Biosystems, Warrington, UK) using a phosphate buffer (adjusted to pH 7.2) and eluted with phosphate buffer, pH 2. The elution peak was measured by UV detection (280 nm). The product peak was integrated and the monoclonal antibody concentration was determined using a standard curve of purified antibody. The protein A intra-assay variation (within a data set obtained from one experiment) has a coefficient of variation (CV) of ~ 3% while the inter-assay variation (from repeated experiments) was ~ 5 %. The CV for individual variables within an assay is calculated by the ratio of standard deviation to the mean while the CV of a statistical model is calculated by the ratio of the root mean squared error (RMSE) to the mean of the dependent variable.

### **2.4.5 Gyroslab™ xP workstation**

The HCP concentration was determined by Gyroslab™ xP (Gyros, Uppsala, Sweden), a high throughput, automated adaptation of an ELISA that uses CDs with highly defined microstructures, to quantify HCP levels. Samples were diluted in Gyros Rexpip AN buffer and together with in-house reagents and buffers are added onto 96-well microtitre plates. Biotin conjugated in-house sheep derived polyclonal antibodies (pAbs) against CHO HCPs are added to the CDs and spun over the streptavidin coated bed column. Samples are then added to the CDs and the HCPs are captured by the streptavidin bead-biotin pAb complex. Alexafluor647 conjugated pAbs against CHO HCPs are then added to the CD, which bind to the captured HCPs. The level of excitation of Alexafluor647 is proportional to the amount of

bound complexes and therefore HCPs. The quantification of HCPs levels is done by a comparison against a standard curve of known quantities of HCPs. The standard curve has a range of 6.1 – 100,000 ng mL<sup>-1</sup>. The assay's lower limit of quantification is 20 ng mL<sup>-1</sup> while the upper limit is 80,000 ng mL<sup>-1</sup>. Harvest samples which often provide HCP results  $> 1 \times 10^6$  ng mL<sup>-1</sup> were diluted accordingly to ensure that the results reported were within the curve range. If the necessary dilution was 1:2, the sample was also analysed at 1:4 and 1:8 dilutions. Back calculated results from these dilutions need to be within the assay's precision of 30 % CV. Within all ambr experiments performed, the assay variability was within 20 % CV.

#### **2.4.6 Protease assay**

The EnzChek<sup>®</sup> Protease Assay Kit (E6638, Molecular Probes, Eugene, OR) was used to assess protease activity within unclarified harvest samples, resulting from fed-batch mammalian cell culture performed using the ambr system (Chapter 4). The assay is fluorescence-based and is capable of detecting metallo-, serine, acid and sulfhydryl proteases. The substrate, BODIPY FL casein, is heavily labelled with pH-insensitive green-fluorescent BODIPY<sup>®</sup> FL dyes, resulting in almost total quenching of the conjugate's fluorescence. The hydrolysis reaction initiated by the proteases found within the unknown samples, releases highly fluorescent BODIPY FL dye-labelled peptides. The fluorescence increase, measured by a microplate reader, is proportional to the protease activity. The dye fluorescence was excited at 485 nm and the emission intensity at 530 nm recorded using a 495 nm cut-off filter on a Molecular Devices Gemini XPS microplate reader.

##### **2.4.6.1 Reagents preparation**

A 1.0 mg/mL stock solution of the BODIPY<sup>®</sup>-casein conjugate detection substrate was prepared by adding 0.2 mL of PBS to one of the vials containing the lyophilized substrate. This solution was then diluted 100-fold in 1X digestion buffer. The reagent was kept in the dark at 4°C. The 1X digestion buffer was prepared by diluting the 20 X digestion buffer (200 mM Tris-HCl, pH 7.8) 20-folds with deionized water. Due to the nature of the unknown samples to be analysed (unclarified harvest samples) a

high dilution is required in order to accurately measure fluorescence increase. This would involve a higher amount of Tris-HCl digestion buffer than the one provided in the commercial kit, therefore the 20X Tris-HCl digestion buffer used in these experiments was prepared in-house using 7.4 g Tris Base and 21.9 g Tris Acid. Comparable results have been seen between trypsin standard curve using in-house digestion buffer and kit buffer (data not shown).

#### **2.4.6.2 Trypsin standard curve**

A solution of 50 mg mL<sup>-1</sup> trypsin (stock 1) was diluted 100-fold in digestion buffer to make a 500 µg mL<sup>-1</sup> solution (stock 2) and then diluted again 50-fold to make a 10,000 ng mL<sup>-1</sup> solution (stock 3). Stock 3 was then used to prepare a dilution series (10,000 ng mL<sup>-1</sup> – 50 ng mL<sup>-1</sup>) with 1X digestion buffer in 1 mL total volume. An equal 100 µL volume of sample (trypsin) and substrate (BODIPY FL casein) were added to a white 96 wells microplate to generate duplicate samples. The plate was sealed, incubated at 40°C and then read every hour until enzyme depletion was achieved for the highest trypsin concentration. A buffer-only control (blank) was also prepared.

#### **2.4.6.3 Sample analysis**

Three 0.5 mL aliquots were taken at harvest from each culture of AMBR 4. The samples were kept at -80 °C until analysed. Once a sample was thawed and analysis it was subsequently discarded. Once thawed, the samples were diluted to 1.0 mL in 1X digestion buffer in order to make up the following dilutions 1/2, 1/5, 1/10, 1/50, 1/100, 1/500, 1/1000 and 1/2000. An equal volume of sample and substrate (100 µL) was added to a white 96 wells microplate in order to generate duplicate samples for each dilution. Controls without the substrate (sample + digestion buffer) were also prepared for each dilution as well as a blank (buffer-only control). The plate was sealed, incubated at 40 °C and then read every hour until enzyme depletion was achieved. An incubation temperature of 40 °C was chosen for a faster reaction time as well as for consistency between plates (room temperature can fluctuate from one day to another). Raw intensity counts increases of the sample compared to the blank

of less than 20 % were considered as negative for protease activity per manufacturer guidelines.

## **2.5 Statistical analysis**

### **2.5.1 Spearman's Rank correlation test**

The strength of the correlations between variables was evaluated by Spearman's rank correlation test which is a non-parametric test used to measure the strength of association between two variables, where the value  $r_s = 1$  means a perfect positive correlation and the value  $r_s = -1$  means a perfect negative correlation. Correlations with p-values less than 0.05 were considered significant. The formula used to calculate Spearman's Rank is shown below.

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (7)$$

where  $d_i$  is the difference in the ranks given to the two variable values and  $n$  is number of variables in each category.

### **2.5.2 Model building**

In terms of model building, a number of different approaches exist. The aim of model selection is to minimize the number of predictors which account for the maximum variance in the dependable variable. All methods involve optimising the model by including all relevant variables and disregarding variables that only contribute a marginal increase in the predictive power of the model. These are forward addition, backward elimination and stepwise regression.

#### **2.5.2.1 Forward addition**

Forward addition starts only with the intercept and then performs "n" regressions with the intercept and each variable one at a time. The variable that contributed the most to the explanation of the response variable is added to the model. The next step is to perform "n-1" regressions with the intercept, the first variable added to the

model and the variable with the second highest significance from the remaining pool. The process is repeated until none of the remaining variables have a significant contribution to the model, given the variables that are already in the model. The main drawback of forward addition is the possibility of one of the included variables to become insignificant with the addition of newer variables.

#### **2.5.2.2 Backward elimination**

Backward elimination starts by constructing a model that includes all the variables. The least significant variable is then dropped from the model and the remaining variables are re-evaluated. The least significant variable is again dropped and the process is repeated until there are no more variables to be eliminated, all remaining variables are significant. Backward elimination also has its own drawbacks. Sometimes variables are dropped that would be significant when added to the final reduced model.

#### **2.5.2.3 Stepwise regression**

Stepwise regression is a mixture of the forward and backward selection techniques. Stepwise regression is a method of model building by adding or removing variables only based on the  $t$ -statistics of their estimated coefficients. Stepwise regression is a modification of the forward selection so that after each step in which a variable was added, all variables in the model are checked to see if they are significant. Stepwise regression has two levels of significance: one for adding variables and one for removing variables. The criterion for adding a variable to the model should be more rigorous than the criterion for keeping a variable in the model so the process does not get into an infinite loop (Howell, 2013). Stepwise selection is an algorithm for picking a “good” (useful) model (Kadane & Lazar, 2004).

#### **2.5.2.4 All possible regression**

All possible regression analysis tests all possible subsets of potential independent variables. If there are “ $n$ ” potential independent variables then there are  $2^n$  distinct

subsets to be tested. All possible regression chooses the best combination of predictors by running regression analyses for all possible predictors. One difficulty is deciding the optimal criteria to use in choosing the “best” model. A variety of criteria for choosing the best possible model exists, including the highest  $R^2$ , the higher predicted  $R^2$ , the lowest root mean square error (RMSE) or Akaike’s Information Criterion (AIC).

### **2.5.3 Model selection**

In statistical data analysis, multiple competing models are often considered. The purpose of model selection is to identify a model that has a balance between the goodness-of-fit of the data, prediction capability and model complexity. Two of the main model selection criteria are the Akaike information criterion (AIC) and Bayesian information criterion (BIC). For both, a lower value is preferred. The fundamental difference between the two criteria lies in the assumptions made. A model with a lower AIC value indicates the model is closer to the true relationship, whilst a model with a lower BIC value indicates the model is more likely to be the true relationship (Dziak et al., 2012; Vrieze, 2012). For ecological publications from 1993 - 2013 that implemented formal methods for multi-model inference, the frequency of using AIC as a model selection criteria was 84% compared to 14% for BIC and 2% for other approaches (Aho et al., 2014). Information-theoretic criteria like AIC has been widely shown as a superior tool for choosing among statistical models as compared to the  $p$ -value (Anderson et al., 2000; Burnham & Anderson, 2002; Gerrodette, 2011). AIC is best suited for finding the best model for predicting new data (Dziak et al., 2012; Murtaugh, 2014). AIC and BIC are simply measures for comparison of models, allowing relative ranking of several models. As a result, AIC and BIC are not able to quantify the goodness of fit or the actual predictive power of the model. The  $R^2$  and  $R^2$  predicted as well as root mean square error (RMSE) from ANOVA as well as residual plots should be used to evaluate the models. In this analysis, the focus was towards selection of models that are able to best predict the responses (titre and HCP). For this reason predicted  $R^2$  was the first main criteria for choosing between potential models. When  $R^2$  and predicted  $R^2$  values were very

similar between several models, then RMSE and AIC were taken into account. For both a smaller value is preferred.

#### **2.5.4 Model assumptions**

Most statistical tests depend on specific assumptions about the variables used in the analysis. When these assumptions are violated, certain inferences and predictions from the regression analysis may be inaccurate and unreliable, resulting in over- or under-estimation of significance or effect size(s). Common ways to detect violations in regression modelling assumptions involve diagnostic plots and residual tests. The multiple linear regression model is based on several assumptions (e.g. residuals are normally distributed with zero mean, are independent, have a constant variance and the mean of the response, at each set of values of the predictor is a linear function of the predictors) (Alexopoulos, 2010; Williams et al., 2013). The assumptions which will be tested in this analysis are the normality and autocorrelation of the residuals and the linearity of the model.

##### **2.5.4.1 Normality of residuals**

To check the normality of residuals, a histogram plot of the residuals can be used. If the residuals are bell-shaped distributed, this implies that the normality assumption is not violated. A common test to check the normality of residuals is the Shapiro-Wilk test. The test is recommended for sample size of less than 50 (Razali and Wah, 2011). The null hypothesis of this test is that the data comes from a normal distribution. Small  $p$ -values indicate that the hypothesis of normality of the residuals should be rejected.

##### **2.5.4.2 Autocorrelation of residuals**

Durbin-Watson test is generally used to test for the presence of autocorrelation in residuals. Autocorrelation means that the adjacent observations are correlated. If the residuals are correlated, then the regression method does not accurately estimate the coefficient's error resulting in predictors being shown as significant when this might not be the case. The Durbin-Watson statistic will be near 2.0 if there is no

autocorrelation. If the statistic is near 0.0, then there is evidence of positive correlation whereas if the statistic is near 4, there is evidence of negative autocorrelation. The  $p$ -value is for the DW statistic under the null hypothesis that there is no autocorrelation among the residuals. If there is no prior reason to believe that the autocorrelation should be positive or negative, then a two-tailed rejection region should be used. Using  $\alpha = .05$ , the null hypothesis of no autocorrelation should be rejected whenever the  $p$ -value  $\leq 0.025$  or  $p$ -value  $\geq 0.975$  (Chatfield, 2014).

#### **2.5.4.3 Linearity of the model**

Nonlinearity is usually most evident in a residuals vs predicted graph, which is a common output of the standard regression. Ideally, the residual versus predicted plot should show a constant band of residuals on either side of the regression line, with a roughly constant variance, showing no trend in the residuals (Williams et al., 2013).

#### **2.5.5 Model validation**

Validation methods generally involves predicting outcomes for a small subset of the data using the remaining observations and then repeating the process for certain number of other subsets.

##### **2.5.5.1 K-fold cross validation**

Cross validation (CV) is a method of assessing the accuracy and validity of a statistical model. CV is the process through which only part of the data (training set) is used to fit the model. The remaining data (test set) is used to test the model. The size of the training and test sets are determined by the number of folds defined in a  $k$ -fold cross validation. The  $k$ -fold cross validation method randomly divides the data into  $k$  subsets. One subset is used as the test set while the remaining subsets are used as the training set. For  $k=10$ , the data is divided into ten subsets, each representing 10% of all the data. A predictive model is trained with 90% of the data and then validated with the remaining 10%. The process is repeated 9 other times so that each subset of the data serves as the test set. The root mean square errors of the  $k$  subsets

are squared and summed to construct the cross validation SSE (squared sum of prediction errors). The resulting SSE is then used to calculate a  $R^2$  k fold value.

## 2.5.6 Definitions

The following statistical parameters have been used to evaluate the resulting cell culture models. Ideally, models with a high  $R^2$  and  $R^2$  predicted as well as with a low root mean squared error (RMSE) will be chosen.

### 2.5.6.1 Pearson's coefficient (r)

Pearson's coefficient (r) also named correlation coefficient, measures the linear relationship between two variables in a sample and is used as an estimation of the correlation in the whole population.

$$r = \frac{\text{Sum of products}_{XY}}{\sqrt{\text{Sum of squares}_X \times \text{Sum of squares}_Y}} \quad (8)$$

$$r = \frac{SP_{XY}}{\sqrt{SS_X \times SS_Y}} \quad (9)$$

where  $SS_X$  refers to the variance of the X scores,  $SS_Y$  refers to the variance of the Y scores and  $SP_{XY}$  refers to the variance shared between X and Y:

$$SS_X = \sum (X - \bar{X})^2 \quad (10)$$

$$SS_Y = \sum (Y - \bar{Y})^2 \quad (11)$$

$$SP_{XY} = \sum (X - \bar{X}) \times (Y - \bar{Y}) \quad (12)$$

In simple linear regression,  $r^2$  represents the proportion of variance in the dependent variable that can be explained by the model. This is calculated by squaring the sample correlation coefficient (r) between the outcomes and their predicted values. If additional terms are included,  $R^2$ , called the coefficient of determination, is

the square of the regression coefficient (R). This coefficient estimates the amount of variance in the dependable variable accounted for by a linear combination of the predictor variables. The remainder represents the proportion that is present in the error. The regression coefficient describes the relationship between the observed and predicted variables. With a perfect correlation of  $R = +1$  or  $-1$ , then  $R^2 = 1$  and all variability in the dependent variable can be explained by the model. When  $R = 0$ , there is no relationship between the predictor(s) and the dependent variable, and none of the variability can be explained (Palmer & Connell, 2009; Hinton, 2014).

### 2.5.6.2 $R^2$ predicted

$R^2$  value is an indication of how well the model fits the experimental data, while the predicted  $R^2$  is an estimation of how well the model predicts a response value (Rouiller et al., 2012). The predicted  $R^2$  is calculated using regression analysis and indicates how well the model can predict responses for new combinations of variables within the  $-1$  to  $+1$  range, evaluated to develop the model (Abu-Absi et al., 2010).  $R^2$  predicted was calculated as follows:

$$R^2 \text{ pred} = [1 - (\text{PRESS}/\text{SP}_{XY})] \quad (13)$$

where PRESS represents the predicted residual sum of squares. PRESS is given by the following formula:

$$\text{PRESS} = \sum_{i=1}^n (Y_{\text{obs},i} - Y_{\text{model},i})^2 \quad (14)$$

Regression models are generally used to predict response variables. The data available can be used to fit the model. Trying to assess how well the model predicts responses, using the same data that was used to fit the model can give over optimistic results. A solution to this would be to leave out an observation, fit the model with the remaining data and then predict the left out response. This leave-one-out-approach is a type of cross-validation, where a part of the data is used to fit the model while the rest is used to assess the model's prediction capability. For more complicated models with several predictors, the challenge of choosing the best model for

prediction can often arise when a large number of possible model choices exist. In this situation, the model with the smallest PRESS statistic should be chosen (Tarpey, 2010; Howell, 2013).

### 2.5.6.3 RMSE

The root mean squared error (RMSE) measures the difference between the predicted and observed values. These differences are also called residuals and the RMSE is used to combine them into a single measurement. The RMSE of a model prediction with respect to the estimated response  $Y_{\text{model}}$  is given by the square root of the mean squared error:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (X_{\text{obs}i} - X_{\text{model},i})^2}{n}} \quad (15)$$

where  $X_{\text{obs}}$  is observed values and  $X_{\text{model}}$  is predicted values at time/place

## 2.6 Software used

### 2.6.1 JMP and Design Expert

JMP and Design Expert have been used to analyse high-throughput data resulting from ambr run cultures. JMP and Design Expert are statistical software from SAS and Stat-Ease, respectively, capable of designing and analysing factorial design experiments. The software helps detect main effects and interactions as well as providing statistical equations and diagnostic plots for evaluation of models. Design Expert is an entry-level program able to perform basic statistics as well as linear regression while in addition JMP is capable of performing more advanced analysis such as multivariate analysis, statistical process control, reliability analysis, non-parametric tests and a wider range of regression analysis as well as advanced graphics. Design Expert is more user-friendly and generally the results are easier to interpret than those from JMP. Certain features from each software have been used in this analysis (Chapter 5). The advanced multivariate analysis from JMP was used to

perform the stepwise all possible regression analysis as well as diagnostic plots and different methods for validating the underlying assumptions of the selected models while Design Expert was used for obtaining the predicted  $R^2$  statistic and the chosen model coefficients.

### **2.6.2 C#, Microsoft Visual Studio and Microsoft Access**

The bioprocess economics and optimisation tool used was developed using the programming language C# (C-sharp) which runs using the .NET framework (Microsoft Visual Studio 2008, Microsoft Corporation, WA, USA). This was linked to a database (Microsoft Access, Microsoft Corporation, WA, USA). The tool, developed in the UCL Decisional Tools team, previously used by Simaria et al. (2012) and Allmendinger et al. (2014a) was updated to incorporate statistical cause-and-effect correlations able to link mAb titre and HCP levels with key cell culture parameters. Cell culture factors such as culture temperature, seeding density, media osmolality, pH and timing of feed initiation were also integrated within the database. The optimisation tool was adjusted to take into account different levels of HCP.

## **Chapter 3**

### **3 Quality by design approach to cell culture process development**

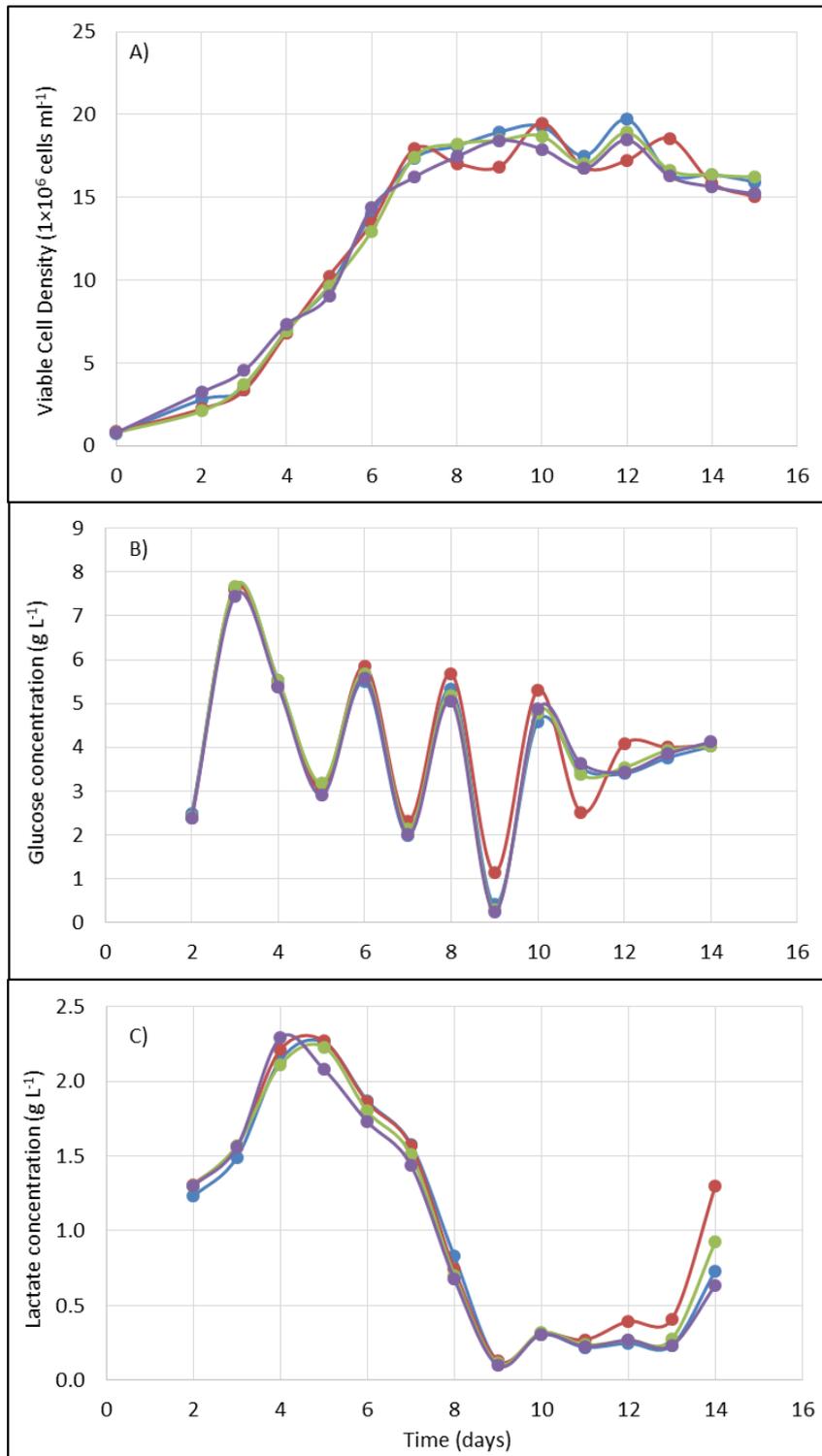
#### **3.1 Introduction**

As the antibody sector has matured, it has seen significant increases in upstream (USP) productivities, with titres reaching over 10 g L<sup>-1</sup> in fed-batch cultures. However, given the complex set of interactions that can affect cell culture performance, it is hard to predict the consequences of titre increase on the host cell protein (HCP) levels at harvest and hence the robustness of downstream operations. Hence it is critical to have systematic methods to explore such interactions and improve process understanding of mammalian cell culture processes.

The aim of this chapter was to use a QbD approach to characterise cell culture performance associated with different generation numbers and explore the consequences of titre increase on the HCP levels at harvest.

#### **3.2 Reproducibility of ambr system**

Within each AMBR experiment, aside from the DoE design, three additional centre points were added to each culture station. This was done in order to evaluate the ambr's ability to provide consistency between vessels with identical conditions. The culture profiles of the four centre points within culture station 1 of AMBR 1 are presented in Fig. 3-1. Good vessel-to-vessel consistency has been observed within ambr vessels cultured under identical fed-batch conditions, in terms of viable cell density, glucose and lactate concentration (Fig. 3-1). Good comparability was also seen in terms of viability, antibody titre (within 10 %) and HCP levels (within 20 %) (data not shown). There was also a good vessel-to-vessel comparability between centre point conditions from AMBR 1, CS2, AMBR 2 and AMBR 3, but only AMBR 1, CS1 was chosen to be shown here.



**Figure 3-1: Fed-batch culture profiles for centre points within AMBR 1 A) viable cell density, B) glucose concentration, C) lactate concentration. The cultures were seeded at  $0.8 \times 10^6$  cells  $\text{ml}^{-1}$  viable cell density, 353  $\text{mOsm kg}^{-1}$  and 36.5 °C.**

### 3.3 Growth profiles of CHO cells in fed-batch culture

AMBR 1, 2 and 3 correspond to low, mid and high generation cells within DoE designs exploring the impact of culture temperature, starting media osmolality, seeding density, culture pH, timing of feed initiation and cell generation number on cell growth, specific antibody productivity, antibody titre and HCP concentration at harvest. The experimental design for AMBR 1, AMBR 2 and AMBR 3 are summarised in Table 3-1. The cell culture profiles for AMBR 1, AMBR 2 and AMBR 3 experiments using different generation number cells and their corresponding titre and viability profiles are summarized in Fig. 3-2, 3-3, 3-4, respectively. The growth profiles within all fed-batch cultures using a CHO high producing cell line showed a normal growth pattern. Within low generation cultures (AMBR 1), a maximum viable cell density (VCD) range of  $13 - 27 \times 10^6$  cells  $\text{mL}^{-1}$ , corresponding to a titre range of  $6 - 8.3$  g  $\text{L}^{-1}$  and a final viability of 55 - 68 % was seen at 36.5 °C, while a maximum VCD range of  $12 - 19 \times 10^6$  cells  $\text{mL}^{-1}$ , corresponding to a titre range of  $6 - 8.2$  g  $\text{L}^{-1}$  and a final viability of 70-84 % was seen at 33 °C. Within mid generation cultures (AMBR 2) a maximum VCD range of  $3 - 18 \times 10^6$  cells  $\text{mL}^{-1}$ , corresponding to a titre range of  $1 - 6.8$  g  $\text{L}^{-1}$  and a final viability of 43 - 75 % was observed. Within high generation cultures (AMBR 3), a maximum VCD range of  $19 - 26 \times 10^6$  cells  $\text{mL}^{-1}$ , corresponding to a titre range of  $6 - 7.4$  g  $\text{L}^{-1}$  and a final viability of 61 - 74 % was seen at 36.5 °C, while a maximum VCD range of  $12 - 21 \times 10^6$  cells  $\text{mL}^{-1}$ , corresponding to a titre range of  $6.8 - 9.6$  g  $\text{L}^{-1}$  and a final viability of 65 - 83 % was seen at 33 °C.

The culture with the highest growth profile within AMBR 1 (Fig. 3-2, A)) was cultivated at 36.5 °C, 313 mOsm  $\text{kg}^{-1}$  media osmolality and  $1 \times 10^6$  cells  $\text{mL}^{-1}$  seeding density, while the lowest growth profile achieved was cultured at 36.5 °C followed by a temperature shift to 33 °C after day 4 of culture, 394 mOsm  $\text{kg}^{-1}$  media osmolality and  $0.5 \times 10^6$  cells  $\text{mL}^{-1}$  seeding density. The culture with the highest growth profile within AMBR 2 (Fig. 3-3, A)) was seeded at a  $0.75 \times 10^6$  cells  $\text{mL}^{-1}$  seeding density and a  $6.8 \pm 0.1$  pH was maintained while the lowest growth profile achieved was seeded at  $0.5 \times 10^6$  cells  $\text{mL}^{-1}$  seeding density and a  $6.6 \pm 0.1$  pH was maintained. The culture with the highest growth profile within AMBR 3 (Fig. 3-3,

A)) was cultivated at 36.5 °C, 320 mOsm kg<sup>-1</sup> media osmolality and 1.25 × 10<sup>6</sup> cells mL<sup>-1</sup> seeding density, while the lowest growth profile achieved was cultured at 36.5 °C followed by a temperature shift to 33 °C after day 4 of culture, 390 mOsm kg<sup>-1</sup> media osmolality and 0.5 × 10<sup>6</sup> cells mL<sup>-1</sup> seeding density. It appears that the different generation number cells used in the three sets of experiments, has the biggest influence on antibody titre at 36.5 °C. This will be discussed further in following sections.

### **3.4 Effect of cell culture inputs and generation number on cell growth**

It has been extensively reported in literature that a decrease in temperature to 28-35 °C from 36.5 - 37 °C (the conventional cultivation temperature for CHO cells) can arrest cell growth and prolong viability (Fox et al., 2004; Tait et al., 2013; Mason et al., 2014). This reduced growth rate has been shown to occur due to a cell cycle arrest in the G1 phase (Trummer et al., 2006), which also causes the cells in G1 phase to be larger in size (Becerra et al., 2012). Moore et al. (1997) explains that low temperatures induce a significant suppression of cell death due to apoptosis while cell viability is preserved in temperature shift cultures, owing to a delay in the onset of apoptosis. Kaufmann et al. (1999) reported that cultures run at 37 °C show a much higher cell density compared to those having undergone a temperature shift. The increase in cell density was accompanied by lower cell viabilities at the end of the culture compared to cultures under hypothermic conditions. The progression of viable cell density and cell viability during fed-batch cultivation is presented in Fig. 3-2, A) and B) for AMBR 1 (low generation), Fig. 3.3 A) and B) for AMBR 2 (mid generation), and Fig. 3-4 A) and B) for AMBR 3 (high generation). In comparison to the control cultures (where temperature was maintained at 36.5 °C), a temperature decrease to 33 °C significantly affected cell growth and viability. Reduced growth was observed at 33 °C while a higher viability was maintained throughout the culture, which agrees with literature (Abu-Absi, 2010). The CCH viable cells for cultures at 33 °C were on average 25 % lower than those achieved for cultures under optimum temperature, while the viability for the lower temperature cultures was on average 30 % higher at harvest point than cultures maintained at 36.5 °C.

Most cell culture media are designed to have an osmolality in the range of 260 - 330 mOsm kg<sup>-1</sup>, so as to be approximately isotonic with human serum at 290 mOsm kg<sup>-1</sup> (Li et al., 2000). It has been widely investigated in literature the effect an increase in media osmolality has on cell growth, specific cell productivity and volumetric productivity. It was found that high osmolality media depressed cell growth and increased  $q_{\text{Mab}}$  (Lin et al., 1999; Ryu et al., 2000; Lee et al., 2003; Li et al., 2010). Lin et al (1999) explains that the increase in osmotic pressure lowers the water retention within the cells, resulting in shrinkage of cell size that may be the cause of cell growth inhibition. Within our experiments, the highest cell growth was achieved using base media (313 mOsm kg<sup>-1</sup>) while the lowest cell growth was achieved using high osmolality media (~ 394 mOsm kg<sup>-1</sup>). At 36.5 °C and low seeding density ( $0.5 \times 10^6$  cells mL<sup>-1</sup>), the maximum VCD was decreased by approximately 26 % and 51 % respectively when media osmolality was increased from 313 to 353 mOsm kg<sup>-1</sup> and from 313 to 394 mOsm kg<sup>-1</sup> within AMBR 1 cultures. At high seeding density ( $1 \times 10^6$  cell/mL), a decrease of 13 % and 19 % in maximum VCD was seen. In temperature shift cultures to 33 °C, the decrease in maximum VCD at low seeding density was 23 % and 39 %, while at high seeding density it was 6 % and 14 % respectively. Cultures seeded at a higher seeding density as well as cultures incubated under hypothermic conditions (33 °C) are less sensitive to increases in osmolality than cultures at 36.5 °C and low seeding density. These findings match those in literature; with Lee et al. (2003) reporting a decrease of 76 % in maximum VCD when media osmolality was increased from 300 to 450 mOsm kg<sup>-1</sup> and Kim et al (2002) presented an 80 % decrease in maximum VCD when osmolality was increased from 294 to 459 mOsm kg<sup>-1</sup>.

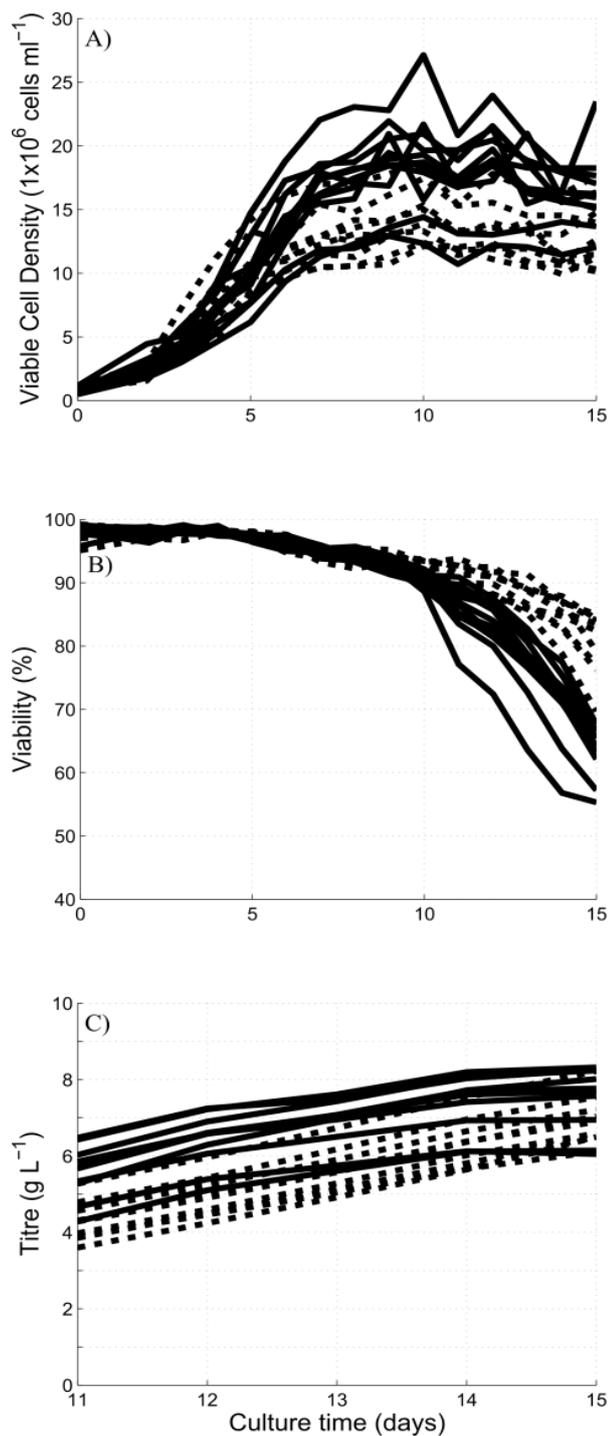
Culture pH is an important environmental factor that should be accurately controlled to ensure the quality of the desired product. A wide range of optimum pH levels have been reported in literature, leading to the conclusion that the optimum pH for growth and recombinant protein production is cell line specific (Trummer et al., 2006). A decrease in culture pH from its optimum has been shown to significantly decrease cell growth (Tsao et al., 2005; Trummer et al., 2006). A slightly acidic condition was found to result in reduced glucose consumption and lactate build-up, leading to an increase in cell death (Tsao et al., 2005). Cultures that were maintained

at a pH of 6.6 showed a significant reduction in growth (by up to 3 fold) (Fig. 3-3, A)) as well as significantly lower specific cell productivity compared to cultures maintained at pH 6.8 and 7. A higher pH is accompanied by a higher cell specific glucose consumption rate, which gives a higher cell growth rate in the exponential phase (Li et al., 2010). Cultures maintained at a pH of 6.8 and 7 did not show a significant difference in the levels of VCD achieved, which shows that for this specific cell line, a pH range of 6.8 - 7 was suitable for cell growth. Link et al. (2004) also found a pH in the range of 6.8 - 7 to give the best growth for a recombinant MUC 1 fusion protein expressed by CHO-K1 cells.

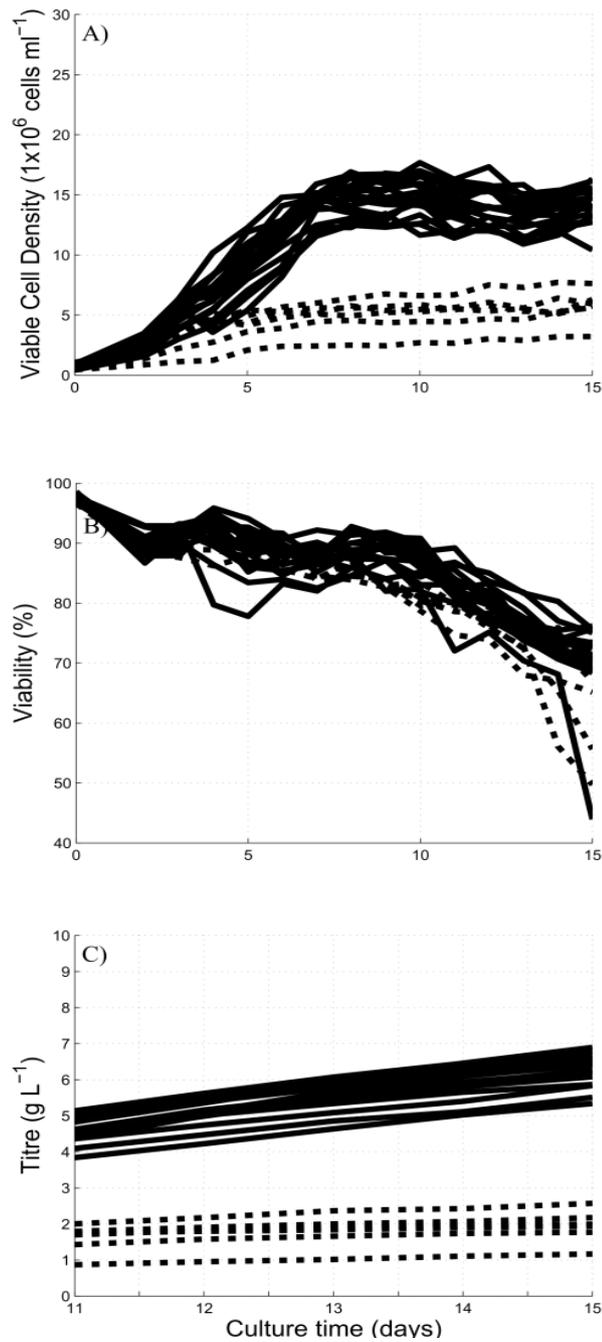
Cell generation number is seen to have an impact on cell growth. Within cell cultures with a higher generation number (generation 44 and 54, compared to generation 33) an increase of 8.7 to 40.4 % in CCH viable cells is observed (Table 3-2).

**Table 3-1: Experimental design setup for AMBR 1, AMBR 2 and AMBR 3 experiments**

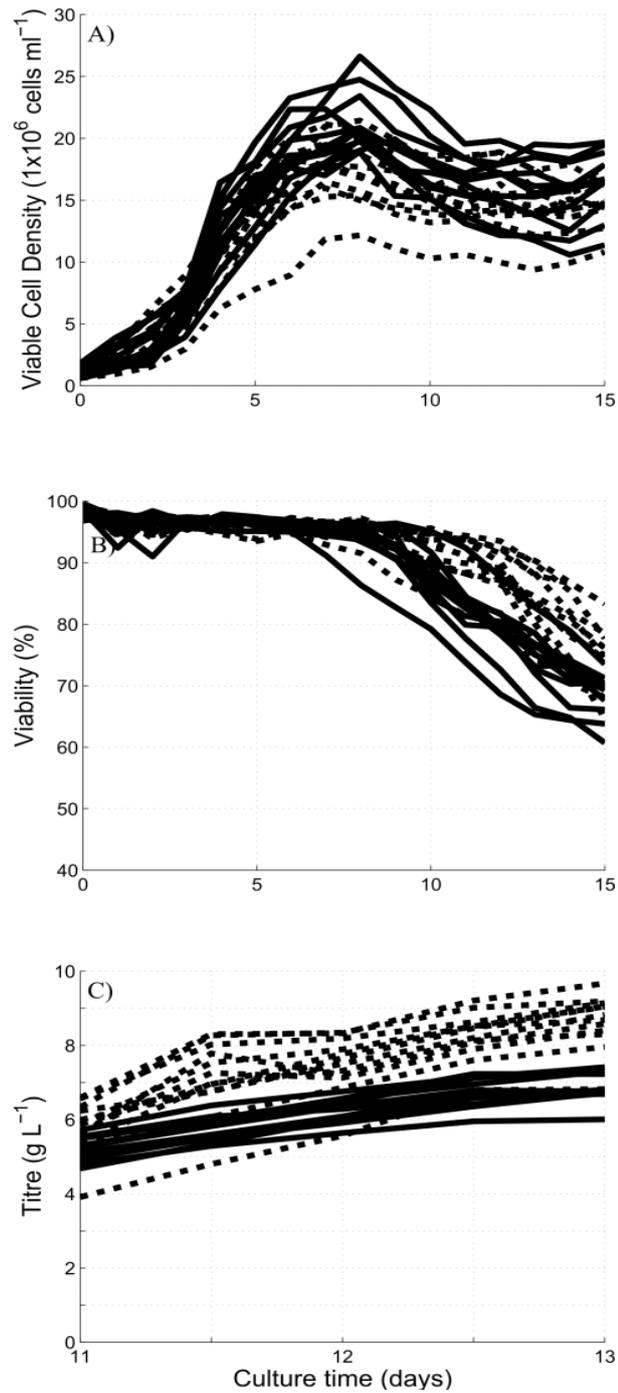
	AMBR 1				AMBR 2				AMBR 3			
	CS1		CS2		CS1		CS2		CS1		CS2	
Number of experiments	12		12		12		12		12		12	
Number of outliers removed	1		3		0		0		1		0	
Generation number	33				44				54			
Seeding density ( $\times 10^6$ cells mL <sup>-1</sup> ) *(expected)	0.49 0.8 1.14	(0.5)* (0.75) (1)	0.59 0.82 1.14	(0.5)* (0.75) (1)	0.42 0.69 1.06	(0.5)* (0.75) (1)	0.9	(1)	0.58 1.08 1.8	(0.5)* (1) (1.65)	0.58 1.19 1.8	(0.5)* (1) (1.65)
Osmolality (mOsm kg <sup>-1</sup> ) *(expected)	314 353 394		314 353 394		313		313		310 331 364	(310)* (350) (390)	319 356 389	(310)* (350) (390)
Temperature (°C)	36.5		33		36.5				36.5		33	
pH	6.8				6.6 6.8 7		6.6 6.8 7		6.8			
Timing of feed initiation (day)	2		2		2		1 2 3		2		2	



**Figure 3-2: Fed-batch culture profiles for AMBR 1 A) viable cell density, B) viability profiles, C) antibody titre. Two different temperatures have been used -- 36.5 °C, ... 33 °C. Within the DoE design, cultures were seeded at a density of 0.49, 0.8, 1.14  $\times 10^6$  cells  $\text{mL}^{-1}$  and the osmolality levels used were 314, 353, 394  $\text{mOsm kg}^{-1}$  respectively. The pH was maintained at  $6.8 \pm 0.1$  and feeding started on day 2 of culture.**



**Figure 3-3: Fed-batch culture profiles for AMBR 2** A) viable cell density, B) viability profiles, C) antibody titre. Three different pH levels have been used  $\cdots$  6.6,  $-$  6.8 and 7. Within CS1 of AMBR 2 cultures, seeding density and pH levels were  $0.42, 0.69, 1.06 \times 10^6$  cells mL<sup>-1</sup> and 6.6, 6.8, 7 while feeding started on day 2 of culture. Within CS2 of AMBR 2 cultures, pH and timing of feed initiation levels were 6.6, 6.8, 7 and day 1, 2, 3 respectively, while the starting seeding density was kept constant at  $1 \times 10^6$  cells mL<sup>-1</sup>. A temperature of 36.5 °C and an osmolality of 313 mOsm kg<sup>-1</sup> were maintained for all 24 cultures.



**Figure 3-4: Fed-batch culture profiles for AMBR 3 A) viable cell density, B) viability profiles, C) antibody titre. Two different temperatures have been used -- 36.5 °C, ... 33 °C. Within the DoE design, cultures were seeded at a density of 0.58, 1.1,  $1.8 \times 10^6$  cells  $\text{mL}^{-1}$  and the osmolality levels used were 320, 350, 390  $\text{mOsm kg}^{-1}$  respectively. The pH was maintained at  $6.8 \pm 0.1$  and feeding started on day 2 of culture.**

### **3.5 Effect of cell culture inputs and generation number on $q_{\text{Mab}}$ , titre and HCP levels**

#### **3.5.1 $q_{\text{Mab}}$**

In the past decades the remarkable improvements in specific and volumetric productivities for mAbs have been achieved through extensive bioprocess engineering research (Wlaschin and Hu, 2006) and careful design/selection methods by the use of high-producers or strong clones (Becerra et al., 2012). Different approaches to try and improve the productivity of mammalian-based processes have focused on the increase in VCD and  $q_{\text{Mab}}$  (Tsao et al., 2005; Becerra et al., 2012). Temperature and media osmolality were chosen as parameters within AMBR 1 (low generation) and AMBR 3 (high generation) experiments due to their history of influencing the specific cell productivity (Kaufmann et al., 1999; Ryu et al., 2000; Lee et al., 2003; Becerra et al., 2012). Temperature is known as an important factor in process optimisation (Fox et al., 2004; Trummer et al., 2006). Lower cell culture temperature is often used to shift cell metabolism towards protein production and essentially improving productivity (Yoon et al., 2003; Tsao et al., 2005; Becerra et al., 2012). Initial studies into the effect of low temperature on  $q_{\text{Mab}}$  in mammalian cells were not encouraging, with  $q_{\text{p}}$  either being unchanged (Chuppa et al., 1997) or decreased (Ryll et al., 2000). Several other studies have shown that  $q_{\text{p}}$  can be significantly enhanced by culturing at temperatures in the 30 - 33 °C range (Kaufmann et al., 1999; Yoon et al., 2003). The effect of low temperatures on  $q_{\text{p}}$  seems to be cell-line dependent, with certain CHO cell lines, achieving higher productivity under mild hypothermia (Fox et al., 2004). Within low generation cultures, a 6 – 24 % increase in  $q_{\text{Mab}}$  was seen in temperature shift cultures to 33 °C compared to cultures maintained at an optimum of 36.5 °C, whereas within high generation cultures, a 13 – 43 % increase in  $q_{\text{Mab}}$  was seen (data not shown). It seems that within high generation cell cultures, a shift to a lower temperature has a higher impact on  $q_{\text{Mab}}$ , than within low generation cultures.

High osmolality media created by addition of sugars or salts (e.g. NaCl) has been known as an economical method of increasing  $q_{\text{Mab}}$  in CHO cultures (Lin et al.,

1999; Ryu et al., 2000; Kim et al., 2002; Lee et al., 2003; Lin et al., 2010). Kim et al. (2002) reported increases in  $q_{\text{Mab}}$  of 390 % when media osmolality was increased from 294 to 459 mOsm  $\text{kg}^{-1}$ , whereas Lee et al. (2003) saw an increase of 139 % in  $q_{\text{Mab}}$  when osmolality was increased from 300 to 450 mOsm  $\text{kg}^{-1}$ . It was observed that an increase in initial media osmolality from 313 to 353 mOsm  $\text{kg}^{-1}$  and from 313 to 394 mOsm  $\text{kg}^{-1}$  through addition of NaCl increased the  $q_{\text{Mab}}$  in all cultures (data not shown)

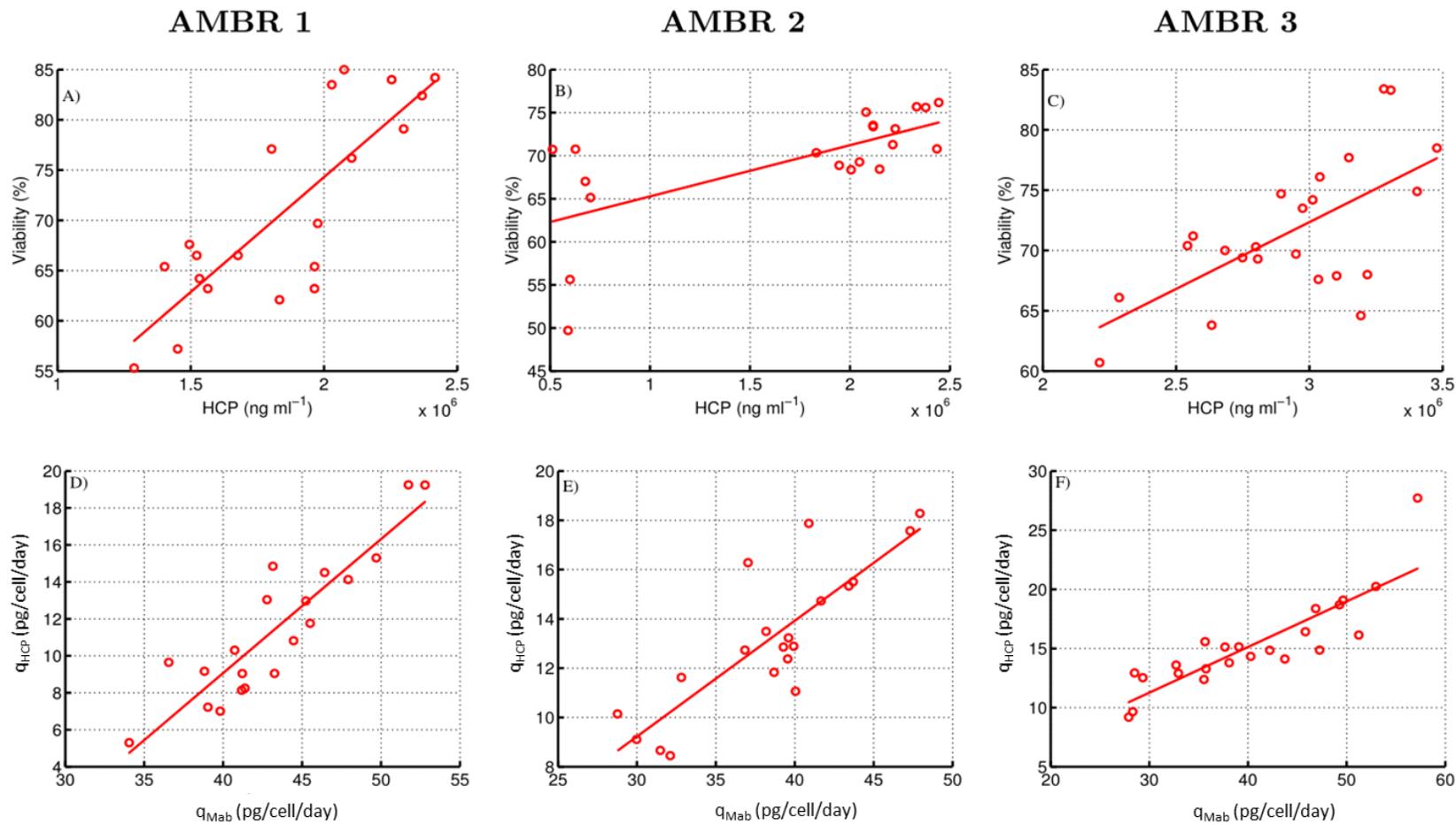
Looking at the relationship between  $q_{\text{Mab}}$  and  $q_{\text{HCP}}$  (Fig. 3-5 D), E), F)), for this cell line the specific productivity for the protein of interest (monoclonal antibody) per cell is always positively correlated to the specific productivity of total HCP per cell, regardless of the generation number used (Table 3-3). As the levels of HCP are used in correlation plots, in this chapter HCP is expressed as  $\text{ng mL}^{-1}$ , as opposed to  $\text{ng mg}^{-1}$  (Chapter 6). This was done in order to identify the true relationship between HCP levels (in absolute value) and antibody titre.

Within AMBR 2 (generation number 44) a 10 % decrease in  $q_{\text{Mab}}$  was observed when compared to lower generation number cultures (AMBR 1 – generation number 33) whereas within AMBR 3 (generation number 54) a 22 % and 16 % decrease in  $q_{\text{Mab}}$ , at 36.5 °C and 33 °C, respectively, was observed, compared to AMBR 1 (Table 3-2). The reduction in  $q_{\text{Mab}}$  with an increase in generation number seen in AMBR 2 and AMBR 3, compared to AMBR 1, is an indication of cell line instability. For cell lines that are considered unstable, a common cause for lower specific productivity for the antibody with an increase in generation number is the loss of recombinant gene copy number (Dorai et al., 2012). The cell line instability problem is further discussed in the next section.

**Table 3-2: The effect of generation number on CCH viable cells,  $q_{\text{Mab}}$ , antibody titre and HCP levels. The comparison is done between replicate conditions of AMBR 2 and AMBR 3 in comparison to AMBR 1.**

	AMBR 1	AMBR 2	AMBR 1	AMBR 3	AMBR 1	AMBR 3	Stability study	
	CS1	CS1 + CS2	CS1		CS2		Mid	Late
Temperature (°C)	36.5		36.5		33		36.5	
Generation number	33	44	33	54	33	54	39	65
CCH viable cells ( $\times 10^6$ cells $\text{mL}^{-1}$ )	3657	3975 (8.7 %)	4688	5414 (15.5 %)	3592	5042 (40.4 %)	3185	3509 (9 %)
$q_{\text{Mab}}$ ( $\text{pg cell}^{-1} \text{ day}^{-1}$ )	42.5	38.2 (-10.1%)	41.0	31.9 (-22.2 %)	47.8	40 (-16.4 %)	11.8	9.4 (-20.8 %)
Titre ( $\text{g L}^{-1}$ )	7.6	6.4 (-15.8 %)	8.2	7 (-14.6 %)	7.0	8.4 (19.5 %)	1.8	1.5 (-16.7 %)
HCP levels ( $\text{ng mL}^{-1}$ )	1,713,760	2,071,358 (20.8 %)	1,820,485	2,414,488 (32.6 %)	2,333,260	3,147,810 (34.9 %)	N/A	

Note: The values presented in this table are averages based on a minimum of two replicates across three AMBR experiments, under similar conditions



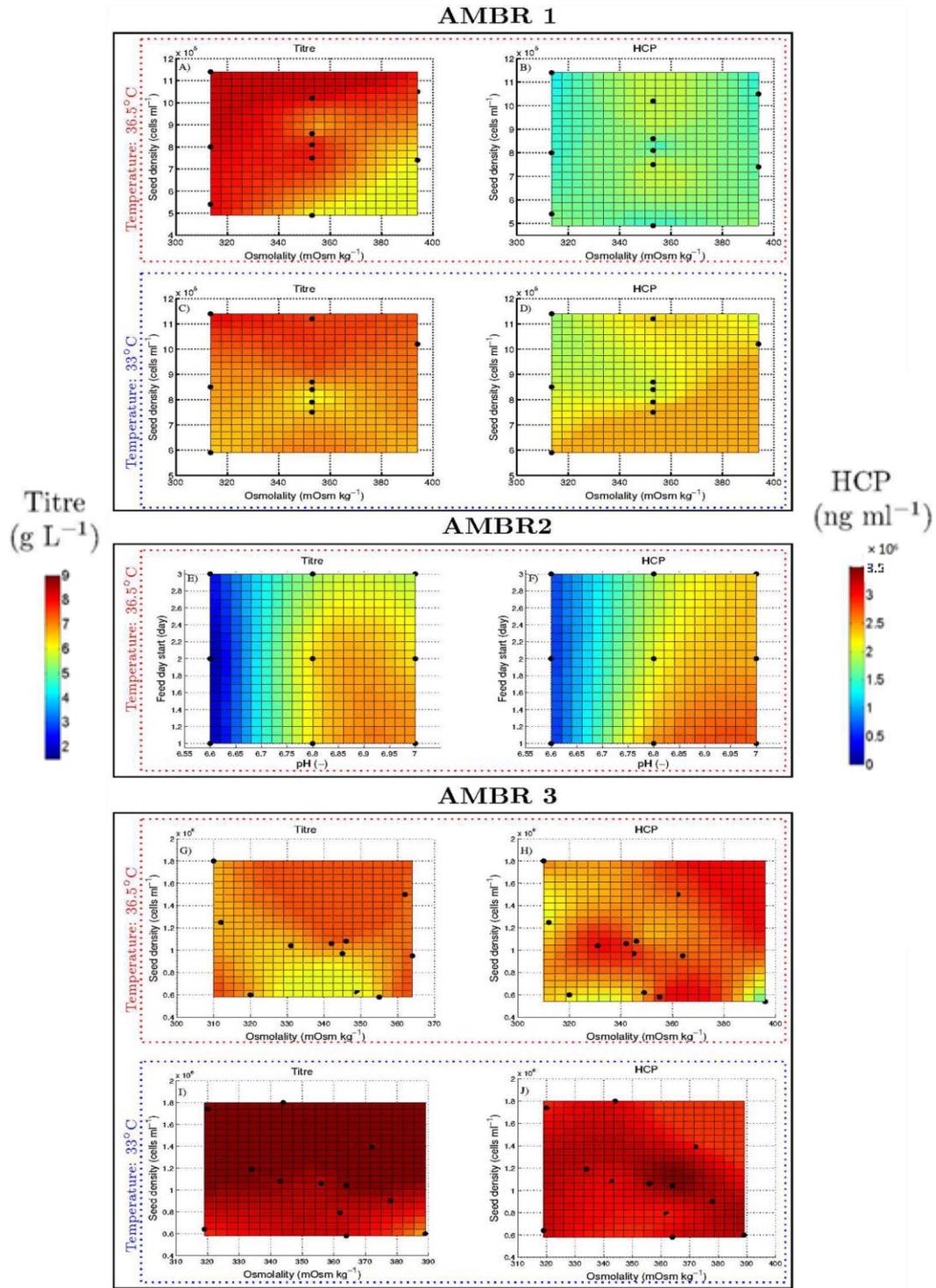
**Figure 3-5: The plot of A), B), C) viability against HCP levels and the observed correlations between D), E), F)  $q_{Mab}$  and  $q_{HCP}$  within AMBR 1, AMBR 2 and AMBR3 experiments, respectively. See Materials and methods for details on AMBR 1, AMBR 2 and AMBR 3 experimental designs.**

### 3.5.2 mAb titre and HCP levels

Antibody titres can be increased through two different methods; increasing the integral area of viable cell days or increasing the cell specific productivity (Kim et al., 2002). Temperature is a key parameter that influences cell growth and recombinant protein production within cell culture, due to its effect on the cell cycle (Yoon et al., 2003). The effect of the different cell culture parameters within the DoE designs on titre and HCP levels is shown using 2D plots in Fig. 3-6 for AMBR 1, 2 and 3 (low, mid and high generation, respectively). Within AMBR 1 cultures, CHO cells cultured at 33 °C achieved overall lower titre levels (average titre of 6.4 g L<sup>-1</sup> across all 12 cultures) compared to cells maintained at 36.5 °C (average titre 7.2 g L<sup>-1</sup>). There was a 5-20 % decrease in titre for cultures at 33 °C, depending on other culture conditions within the DoE design. The correlation between antibody titre, viable cell density, antibody productivity, HCP levels and the  $q_{\text{Mab}}:q_{\text{HCP}}$  ratio is shown in Fig. 3-7 for AMBR 1, 2 and 3. The increase in titre within AMBR 1 was correlated with an increase in viable cell density (Fig.3-7 A)), therefore the increase in  $q_{\text{Mab}}$  which occurred in 33 °C cultures did not compensate for the lower cell growth within these cultures, resulting in lower titres compared to cultures at 36.5 °C. These findings are in agreement with literature, where Trummer et al. (2006) reported a 1.5 fold decrease in titre levels in cultures run at 33 °C compared to cultures at 37 °C and Abu-Absi et al. (2010) also reported decreased product titre at a lower production bioreactor temperature. It should be noted that in both cases, the lower culture temperature was set and maintained from the start of the culture.

The increase in titre within AMBR 3 cultures was correlated with an increase in  $q_{\text{Mab}}$  (Fig.3-7 C)) and in this scenario, higher titres are seen in 33 °C cultures (average 8.6 g L<sup>-1</sup>), condition that favours higher  $q_{\text{Mab}}$ , compared to cultures maintained at 36.5 °C (average titre 6.9 g L<sup>-1</sup>). There was a 15-30 % increase in titre for cultures at 33 °C, depending on other culture conditions within the DoE design. Similar results have been reported by Rameez et al. (2013) which found that a two fold increase in titre occurs in temperature shift cultures, performed in the ambr system. In comparison to our experiments in which the temperature was shifted to 33 °C in day 4 of culture, Rameez et al. (2013) performed the temperature shift from 37

°C to 33 °C, in day 8 of culture, in order to allow maximal viable cell density to be achieved.



**Figure 3-6: Impact of different cell culture parameters (seeding density, media osmolality and timing of feed initiation) on antibody titre and HCP levels at two different temperatures ... 36.5 °C, ... 33 °C, within AMBR 1, AMBR 2 and AMBR 3 experiments. See Materials and methods for details on AMBR 1, AMBR 2 and AMBR 3 experimental designs.**

The effect of temperature on HCP levels within AMBR1 and AMBR 3 cultures is shown in Fig. 3-6. In both set of experiments, a higher level of HCPs was seen in 33 °C cultures compared to 36.5 °C. A 25 % and 16 % increase in HCP levels in 33 °C cultures was seen on average for AMBR 1 and AMBR 3 experiments, respectively. This is in agreement with Tait et al. (2013) which has seen the HCP content determined by an industry standard ELISA assay being ~ 50 % higher in the culture grown under mild hypothermic conditions (32 °C) compared to cultures grown at 37 °C. Cultures at 33 °C are also known to have higher viabilities at harvest compared to cultures at 36.5 °C (Fig. 3-2 B), Fig. 3.4 B)) (Trummer et al., 2006; Abu-Absi, 2010; Rameez et al., 2013). This explains why within all experiments using different cell generations, HCP levels show a positive correlation with culture viability at harvest (Fig. 3-5, A), B), C)). The same behaviour was seen by Grzeskowiak et al. (2009) which observed higher HCP levels in higher viability cultures using HCP ELISA. Yuk et al. (2015) also shows that high density CHO cultures ( $>10^7$  cells mL<sup>-1</sup>) operated in fed-batch mode and having high viabilities (>70 %) throughout the culture duration, can accumulate a high amount of immunogenic HCP (1-2 g L<sup>-1</sup>) in the extracellular environment at the time of harvest (day 14). LDH results showed that cumulative cell lysis can be considerable in high density fed-batch CHO cultures, despite high viability readings. These publications offer a possible hypothesis for why this might be happening but for this set of experiments, we cannot say with confidence that this applies here.

An increase in media osmolality was shown to enhance  $q_{\text{Mab}}$  (Lin et al., 1999; Kim et al., 2002; Lee et al., 2003; Lin et al., 2010), however the use of high osmolality media did not increase the final antibody concentration due to cell growth being depressed at elevated osmolality (Kim et al., 2002; Lee et al., 2003). Fig 3-2 shows the impact of media osmolality increase on titre and HCP levels at 36.5 °C and 33 °C. As an increase in media osmolality occurs, a significant decrease in titre levels is seen in all cultures seeded at low cell density. The negative effect of increased media osmolality starts to decline, with an increase in seeding density. Cultures seeded at a high cell density are not sensitive to changes in osmolality, within the osmolality ranges investigated.

To be able to take advantage of the increased  $q_{\text{Mab}}$  in high osmolality cultures in order to increase the final antibody concentration, several strategies have been proposed in literature: 1) biphasic culture, where the osmolality was maintained at a physiological level to promote growth and then increased to maximize  $q_{\text{Mab}}$  (Kim et al., 2002), 2) addition of osmoprotective compounds to the cultures (e.g. glycine betaine) which can improve the cell growth at elevated osmolality, sustaining the enhanced  $q_{\text{Mab}}$  (Ryu et al., 2000).

Within AMBR 1 and AMBR 3 cultures, the increase in media osmolality does not have a significant impact on HCP levels (Fig. 3-6). I am not aware that the effect of osmolality on CHO HCP levels at harvest point has been previously reported in the literature.

Culture pH is a parameter known to significantly influence cell growth and recombinant protein production (Link et al., 2004; Yoon et al., 2005; Trummer et al., 2006; Seo et al., 2013). Significantly lower antibody titres were achieved in cultures at pH 6.6 as compared to cultures at pH 6.8 and 7 (Fig. 3-3 C)). This is mainly due to the significant decrease in viable cell numbers in cultures maintained at pH 6.6 (Fig. 3-3 A)). Jardon and Garnier (2003) studied the effects of pH in the ranges 6.7 – 7.7 for the production of recombinant adenovirus vector (rAV) using HEK293 cells and found that significantly lower titres were seen in cultures at pH 6.7 as compared to pH 7.2. I am not aware that the effect of pH on CHO HCP levels at harvest point has been previously reported in the literature. Fig 3-2 shows the effect of culture pH on HCP levels within AMBR 2 cultures. It can be seen that HCP levels for all cultures at pH 6.6 are significantly lower compared to cultures maintained at pH 6.8 and 7 which correlates with significant lower growth in these cultures (Fig. 3-3 A)). Timing of feed initiation was not found to have a significant impact on antibody titre and HCP levels at low pH. At pH 6.8-7, higher titres are seen in cultures for which feeding started on day 1 of culture, but this correlates with an equivalent increase in HCP levels.

In terms of biopharmaceutical production, cell line stability can be defined as reproducible product concentration and quality for a given cell line over prolonged periods of time, starting from thawing the master cell bank to a time point

determined to be the longest duration allowed for production (Wurm, 2015). At 36.5 °C, within higher generation cell cultures (AMBR 2 – Gen 44; AMBR 3 – Gen 54) an increase in CCH viable cells, a decrease in  $q_{\text{Mab}}$  and a decrease in antibody titre can be observed. The comparison is done between replicate conditions of AMBR 2 and AMBR 3 in comparison to AMBR 1. The trends are consistent with a previous stability study which shows a stability issue as described in Table 3-2. Other process conditions within AMBR 1 and AMBR 3, at 36.5 °C were the same with the previous stability study. There are several studies in literature that have presented on the instability of protein production from recombinant cell lines. The most common causes were identified as a loss of recombinant gene copy number as well as the appearance of a secondary, less productive population of cells (Dorai et al., 2012). Comparing AMBR 3 with AMBR 1 at 33 °C cultures, on average, CCH viable cells increased,  $q_{\text{Mab}}$  decreased and antibody titre increased between replicate conditions.

In literature, a reduction in culture temperature from optimum (36.5 °C) to 33 °C is known to promote an increase in  $q_{\text{Mab}}$  whereas a lower cell growth is generally observed, trends which are more likely to mask a cell line stability issue compared to cultures maintained at 36.5 °C. At higher generation numbers (AMBR 2, 3) compared to AMBR 1, a higher HCP level was seen in all cultures, regardless of culture temperature or other cell culture inputs.

### **3.6 Analysis**

Overall monoclonal antibodies titres can be increased or decreased due to a combination effects from changes in cell densities and/or specific cell productivities (Li et al., 2010), each with a different impact on impurity profiles. This depends on factors such as the cell line characteristics, bioreactor operating parameters, media or feed type and the feeding strategy. These have a significant influence on cell growth, productivity, viability, impurity profiles and may also have an impact on product quality. Increases in cell densities can be achieved by media optimisation that has a great impact on maximizing and maintaining viable biomass, feeding strategies and optimal control of culture conditions such as pH, DO and glucose (Fike, 2009). It is normally anticipated that when high titres are achieved, an increase in HCP levels

would also be seen. It was demonstrated that by using a QbD approach to cell culture process development (HT ambr experimentation, HT analytics and DoE) bioreactor scenarios where the ratio of mAb to HCP can be varied, are identified. This can potentially impact the purification strategy for mAbs where HCP contamination is an issue. This opens the potential of designing a bioreactor process in which mAb harvest titre is high whilst reducing HCP levels.

### **3.6.1 Analysis of low generation cell experiments (AMBR 1)**

The experimental design of AMBR 1 is shown in Table 3-1. It was demonstrated that within low generation cultures, conditions known to increase the specific mAb productivity (hypothermic conditions and high osmolality) had a negative impact on titre, mainly due to depressed cell growth at these conditions. Fig. 3-7 A) shows how within AMBR 1 cultures, the increase in antibody titre correlates to an increase in viable cell number (expressed by CCH viable cells) while  $q_{\text{Mab}}$  is decreased. A Spearman's rank correlation test was performed to assess the strength of the correlations between the different variables within AMBR 1, AMBR 2 and AMBR 3 experiments. A summary of the results is presented in Table 3-4. It can be hypothesized that even though  $q_{\text{Mab}}$  is decreasing, the high number of viable cells achieved within the culture is the main driver for the increase in antibody titre. Assessing the relationship between  $q_{\text{Mab}}$  and  $q_{\text{HCP}}$ , as a function of CCH increase (Fig. 3-7 G)), it is observed that as the viable cell density increases, the ratio of  $q_{\text{Mab}}:q_{\text{HCP}}$  is increasing, suggesting that  $q_{\text{Mab}}$  is increasing at a higher rate than  $q_{\text{HCP}}$ , resulting in a purer product which eases the load on the DSP. In this scenario, titre is higher at conditions that correlate with a high number of viable cells within the culture (in these experiments,  $T=36.5\text{ }^{\circ}\text{C}$ , a high seeding density ( $1 \times 10^6\text{ cells mL}^{-1}$ ) and base media ( $313\text{ mOsm kg}^{-1}$ ), as well as an optimum combination of seeding density and starting media osmolality (e.g. medium seed ( $0.75 \times 10^6\text{ cells mL}^{-1}$ ) and medium osmolality ( $353\text{ mOsm kg}^{-1}$ ); high seed ( $1 \times 10^6\text{ cells mL}^{-1}$ ) and high osmolality ( $394\text{ mOsm kg}^{-1}$ )). So overall these observations suggest that the amount of product increases but not at the expense of an increase in HCP levels (Fig. 3-7 D)). Similar plots have been done in order to assess the correlation between antibody titre,  $q_{\text{Mab}}$ ,  $q_{\text{Mab}}:q_{\text{HCP}}$  ratio and growth rate (as opposed to CCH viable cells). The

same conclusion can be drawn, therefore only correlation plots using CCH viable cells are shown. This is consistent with a similar trend presented by Jin et al. (2010), although they saw significant differences in product titre; no major change was seen in the levels and composition of HCPs.

### **3.6.2 Analysis of mid generation cell experiments (AMBR 2)**

The experimental design of AMBR 2 is shown in Table 3-1. Switching to a different design space than AMBR 1 and using a different cell stock, a different scenario is identified, where the increase in antibody titre is correlated with an increase in viable cell density as well as specific mAb productivity (Fig. 3-7 B)). Evaluating the impact of antibody titre increase on HCP levels, as a function of increases in viable cell density, it can be observed that titre and HCP levels increase at a similar rate (Fig. 3-7 E)), therefore the ratio of  $q_{\text{Mab}}: q_{\text{HCP}}$  remains constant over the whole range of CCH viable cells (Fig. 3-7 H)). The influence of culture pH on cell growth is the reason for data clustering in Fig. 3-7 B), E), H). Cultures at pH 6.6 revealed significantly lower growth profiles, compared to cultures at pH 6.8 and 7 (Fig. 3-3 A)).

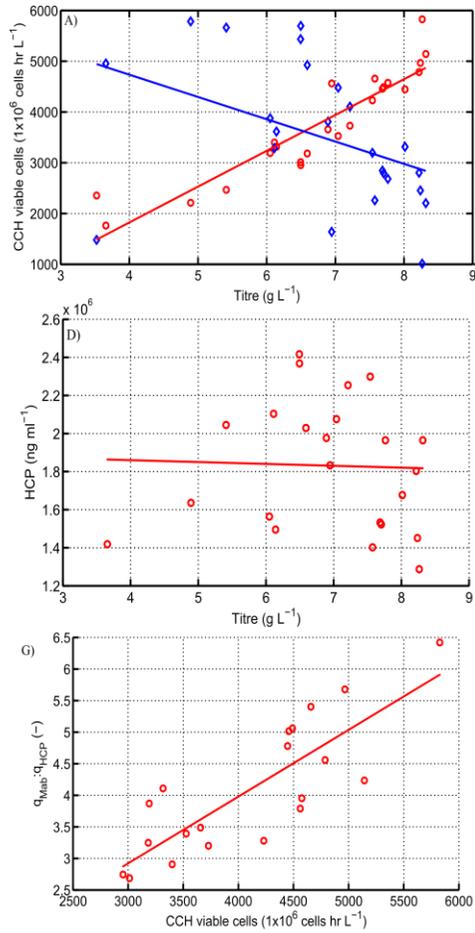
### **3.6.3 Analysis of high generation cell experiments (AMBR 3)**

The experimental design of AMBR 3 is shown in Table 3-1. Extending the design space used for AMBR 1 experiments and using high generation cells, a third distinctive scenario is identified, where the increase in antibody titre is correlated with an increase in  $q_{\text{Mab}}$ , while a constant level of viable cells is seen within all cultures (Fig. 3-7 C)). Similar to AMBR 2, in this case, titre and HCP levels are seen to increase at a similar rate (Fig. 3-7 F)), with the ratio of  $q_{\text{Mab}}: q_{\text{HCP}}$  remaining constant (Fig. 3-7 I)).

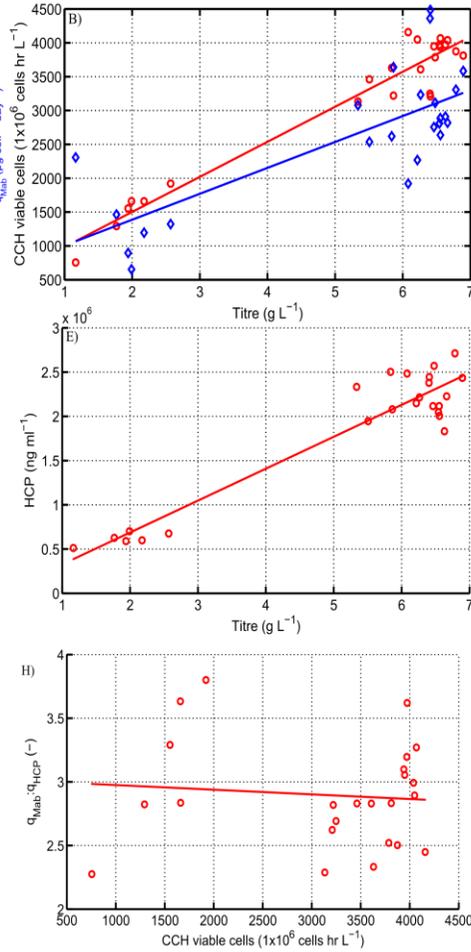
So overall for this cell line, low generation cell number leads to an increase in antibody titre and reduced HCP levels while a mid and high generation number also leads to an increase in antibody titre but with a concomitant increase in HCP level. As mentioned in section 3.5.2, in cell lines that show instability, the appearance of a secondary, less productive population of cells has been identified (Dorai et al., 2012). In the experiments presented in this thesis, this can be potentially reflected by the

higher number of cells (CCH viable cells) and lower specific productivity ( $q_{\text{Mab}}$ ) present in higher generation cultures compared to low generation cultures (Table 3-2). It can be hypothesized that the increase in HCP levels with an increase in antibody titre in mid (AMBR 2) and high generation number (AMBR 3) cultures is attributed to the higher number of cells present in higher generation cultures (showing lower productivity for the mAb).

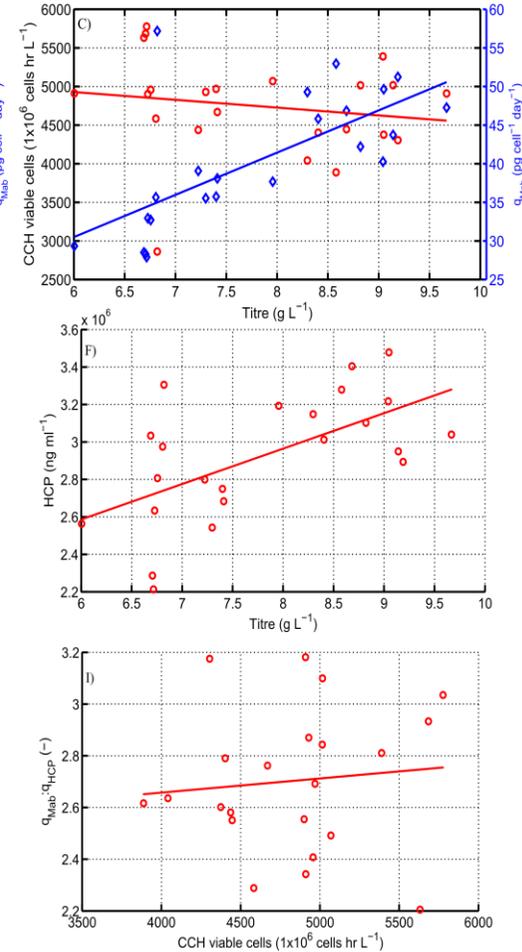
### AMBR 1



### AMBR 2



### AMBR 3



**Figure 3-7: A correlation plot of antibody titre against HCP levels within AMBR 1, AMBR 2 and AMBR 3 experiments. The relationship between A) antibody titre, CCH and  $q_{\text{Mab}}$ , B) antibody titre and HCP levels, C) CCH viable cells and  $q_{\text{Mab}}$ :  $q_{\text{HCP}}$  ratio is shown. See Materials and methods for details on AMBR 1, AMBR 2 and AMBR 3 experimental designs.**

**Table 3-3: Summary of Spearman's rank test for the correlations between variables in Figure 3-5.**

	Variable correlation	Spearman's coefficient ( $r_s$ )	$p$ -value	Summary
	AMBR 1 – low generation (33)			
Figure 3-5 A)	HCP vs Viability	0.789	$4.822 \times 10^{-5}$	Strong positive correlation (significant)
Figure 3-5 D)	$q_{Mab}$ vs $q_{HCP}$	0.83	$8.921 \times 10^{-7}$	Very strong positive correlation (significant)
	AMBR 2 – mid generation (44)			
Figure 3-5 B)	HCP vs Viability	0.299	0.02892	Weak positive correlation (insignificant)
Figure 3-5 E)	$q_{Mab}$ vs $q_{HCP}$	0.788	$5.167 \times 10^{-5}$	Strong positive correlation (significant)
	AMBR 3 – high generation (54)			
Figure 3-5 C)	HCP vs Viability	0.539	0.0078	Moderate positive correlation (significant)
Figure 3-5 F)	$q_{Mab}$ vs $q_{HCP}$	0.893	$3.145 \times 10^{-6}$	Very strong positive correlation (significant)

**Table 3-4: Summary of Spearman's rank test for the correlations between variables in Figure 3-7.**

	Variable correlation	Spearman's coefficient ( $r_s$ )	$p$ -value	Summary
AMBR 1 – low generation (33)				
Figure 3-7 A)	Titre vs CCH	0.890	$2.24 \times 10^{-6}$	Very strong positive correlation (significant)
Figure 3-7 A)	Titre vs $q_{Mab}$	-0.689	0.00079	Strong negative correlation (significant)
Figure 3-7 B)	Titre vs HCP	-0.444	0.05009	Moderate negative correlation (insignificant)
Figure 3-7 C)	CCH vs $q_{Mab}$ : $q_{HCP}$	0.789	$3 \times 10^{-5}$	Strong positive correlation (significant)
AMBR 2 – mid generation (44)				
Figure 3- 7 D)	Titre vs CCH	0.773	$1.52 \times 10^{-5}$	Strong positive correlation (significant)
Figure 3-7 D)	Titre vs $q_{Mab}$	0.644	0.00069	Strong positive correlation (significant)
Figure 3-7 E)	Titre vs HCP	0.578	0.00308	Moderate positive correlation (significant)
Figure 3-7 F)	CCH vs $q_{Mab}$ : $q_{HCP}$	0.489	0.0162	Moderate positive correlation (insignificant)
AMBR 3 – high generation (54)				
Figure 3-7 G)	Titre vs CCH	-0.305	0.156	Weak negative correlation (insignificant)
Figure 3-7 G)	Titre vs $q_{Mab}$	0.775	$1 \times 10^{-5}$	Strong positive correlation (significant)
Figure 3-7 H)	Titre vs HCP	0.571	0.0044	Moderate positive correlation (significant)
Figure 3-7 I)	CCH vs $q_{Mab}$ : $q_{HCP}$	0.230	0.289	Weak positive correlation (insignificant)

### 3.7 Conclusions

Three AMBR experiments, using low, mid and high generation cells were performed in order to evaluate the impact that culture temperature, starting media osmolality, seeding density, culture pH, timing of feed initiation and cell generation number has on cell growth, specific antibody productivity, antibody titre and HCP concentration at harvest. A QbD approach is then presented to help establish the relationship between mAb titre and HCP levels within small-scale fed-batch mammalian cell culture. The results show that a temperature of 36.5 °C is optimum for increased mAb titre, cell growth and results in a lower HCP level as compared to 33 °C. High osmolality media increased  $q_{\text{Mab}}$  but had a negative effect on growth and antibody titre while a pH of 6.6 had a significant negative impact on cell growth, resulting in lower titres and HCP levels within these cultures. A stability issue is observed within higher cell generation cultures at 36.5 °C while at a lower temperature of 33 °C cell line instability is not apparent. Cells of a higher generation lead to an elevated ratio of HCP to product (AMBR 2, 3) as compared to lower generation cells (AMBR 1).

This chapter presents a QbD approach to cell culture process development (HT ambr experimentation and Gyrolab analytics linked with DoE) to show that there is scope for cell culture processes in which the ratio of mAb to HCP can be increased and the association of mAb titre to HCP reduced. It is therefore viable to identify conditions whereby it is possible to increase antibody titre with little impact on HCP levels and hence subsequent DSP operations. (36.5 °C, 313 mOsm kg<sup>-1</sup> media osmolality,  $1 \times 10^6$  cells mL<sup>-1</sup> seeding density, pH 6.8 and low cell generation number).

## **Chapter 4**

### **4 Analysis of problematic HCPs in cell culture processes**

#### **4.1 Introduction**

Recent literature has emphasized the need for a better understanding of the impurity profile entering DSP and the cell culture factors that are likely to change this profile, as certain problematic HCPs have been identified in the final drug product (Aboulaich et al., 2014; Thompson et al., 2014). There is a limited understanding regarding the relationship between the protein of interest, bioprocess conditions (both in USP and DSP) and problematic HCPs (e.g. protease) (Bracewell and Smales, 2013).

In addition to assessing the impact of cell culture inputs on the resulting mAb titre and HCP levels at harvest (Chapter 3), a deeper understanding was provided in this chapter regarding the USP factors influencing the activity of specific problematic HCPs (proteases). The aim of this chapter was to evaluate the impact of several cell culture parameters (temperature, media osmolality and seeding density) on protease activity at harvest as well as to examine the relationship between HCP levels, mAb concentration and protease activity resulting from unclarified harvest samples. A commercially available protease assay was optimised in order to make it suitable for the analysis of unclarified cell culture harvest.

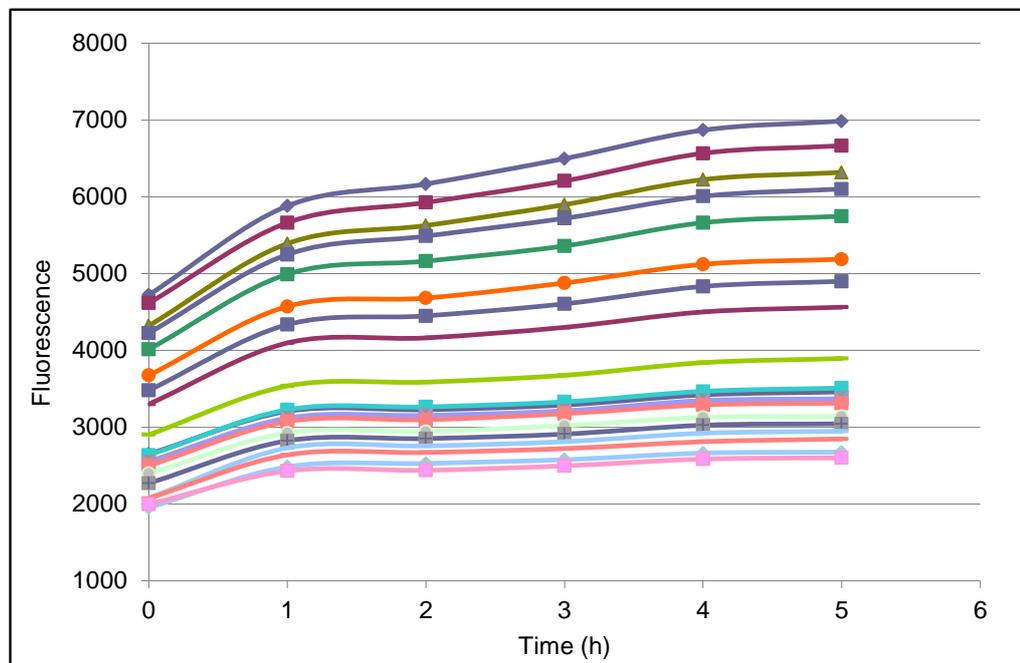
#### **4.2 Protease assay development**

##### **4.2.1 Trypsin standard curve**

The EnzChek<sup>®</sup> Protease Assay Kit (E6638, Molecular Probes, Eugene, OR) was used to measure protease activity within unclarified harvest samples, resulting from fed-batch mammalian cell culture performed using the ambr system. The same protease assay kit was used by Robert et al. (2009) and Dorai et al. (2011) to measure the

amount of protease (aspartic acid protease and serine-threonine protease, respectively).

To be able to determine if the protease assay is capable of measuring protease activity as well as to distinguish between low and high protease activity, a trypsin standard curve was performed, with trypsin concentration ranging from 50 – 10,000 ng mL<sup>-1</sup>. An appropriate enzyme standard of known specific activity that closely matches the protease activity being determined was used. Trypsin is a serine protease that is commonly used in generating standard curves for protease activity assays. Fig 4-1 shows the assay's capability of distinguishing between different concentrations of trypsin; the trypsin concentration is proportional to the fluorescence increase.



**Figure 4-1: Trypsin standard curve. The trypsin concentration varied from 50 – 10,000 ng mL<sup>-1</sup>, in-house Tris-HCl digestion buffer was used for making up the dilutions and the plate was incubated in the dark at 40° C and measured at hourly intervals. The increase in trypsin concentration is proportional to the increase in fluorescence**

#### 4.2.2 Investigate the colour effect of unknown samples

The unknown samples that were used to determine the associated fluorescence increase due to protease activity are unclarified harvest samples resulting from previously performed ambr experiments. A relatively high protease activity was

expected to be present in these samples due to their associated high HCP content and their level of clarification. For this reason, a wide range of dilutions was used in order to determine the appropriate one for measuring protease activity (neat, 1/2, 1/5, 1/10, 1/50, 1/100, 1/500, 1/1000 and 1/2000).

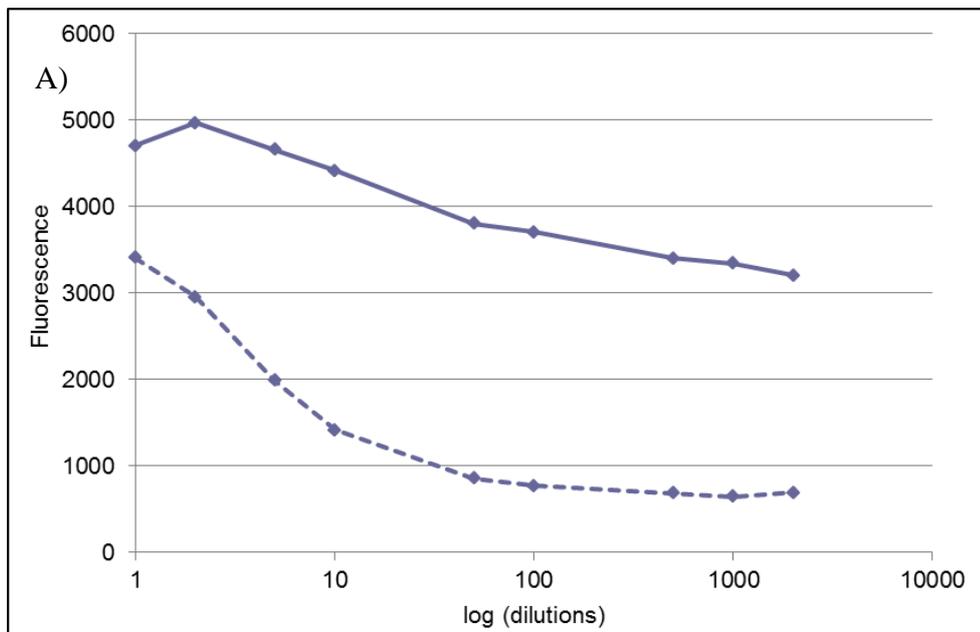
The experimental protocol specifies that an equal volume of sample and substrate (100  $\mu\text{L}$ ) to be added to a 96 wells microplate in order to measure the fluorescence emitted from the sample, which in turn is proportional to the protease activity. The colour of the 1:1 mixture of unknown sample and BODIPY casein substrate (U+S) was quite intense, especially for samples with a low dilution. Coloured samples can give a fluorescence reading, regardless of the presence of protease within the samples. This would make it difficult to assess if the fluorescence increase given by the samples with substrate over time was due to protease activity or due to the intensity of colour within these samples. To verify if coloured samples exhibit a fluorescence increase, a sample of casein substrate (light pink colour), in a 1:1 ratio to digestion buffer (colourless) (S + B) was analysed in order to identify if there is any fluorescence increase due to colour alone. A fluorescence increase of  $\sim 3\text{X}$  higher than the blank (digestion buffer) was observed (data not shown). To identify the fluorescence increase only due to the substrate, the fluorescence background of the buffer was subtracted from the fluorescence background of the substrate and buffer 1:1 mixture:  $S = (B+S) - B$ .

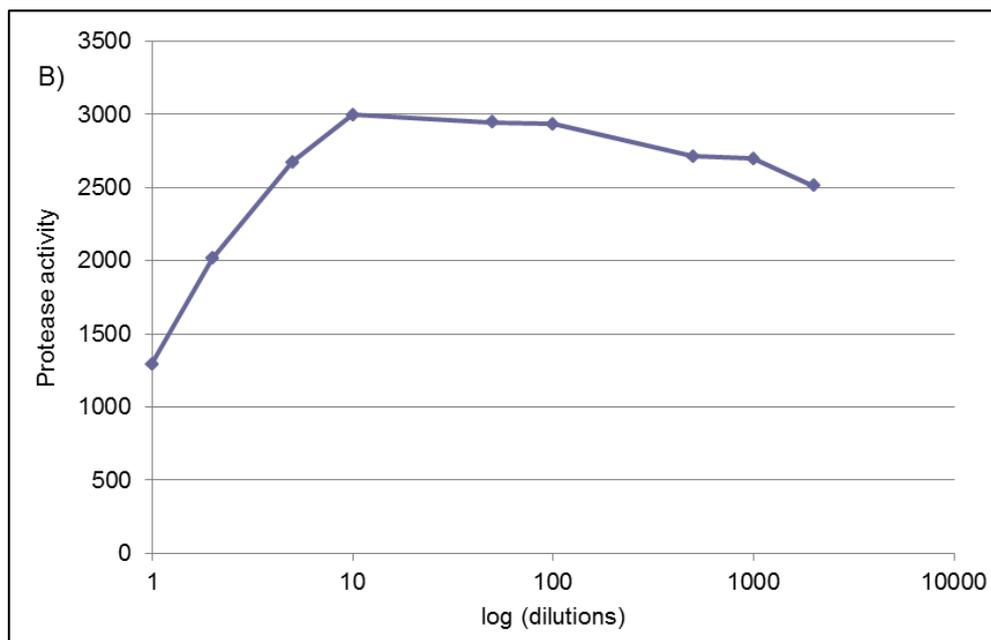
In order to determine if the colour difference between different dilutions had an impact on the fluorescence increase, a fixed concentration of trypsin ( $3000 \text{ ng mL}^{-1}$ ) was incubated with casein substrate for 4 h in a 1:1 dilution. Fig. 4-1 shows that 4 h is sufficient time for most of the  $3000 \text{ ng mL}^{-1}$  trypsin to become depleted. The  $3000 \text{ ng mL}^{-1}$  trypsin in 1:1 mixture with casein substrate was then added to the different dilutions of an unknown sample and the corresponding fluorescence increase was read at time  $t = 0$  (Fig. 4-2 A -). Control samples (without the BODIPY casein substrate) were prepared for each dilution of the sample, as a 1:1 ratio of unknown sample and digestion buffer (U+B) (Fig. 4-2 A...). The true fluorescence increase due to protease activity was then obtained by subtracting the fluorescence background of the substrate-free control (unknown (U) + buffer (B)) as

well as the fluorescence background of the substrate (S) from the unknown sample containing the substrate (unknown (U) + substrate (S)) (Fig. 4-2 B)).

$$\text{Protease activity} = (U+S) - (U+B) - S \quad (16)$$

The background fluorescence given by the casein substrate (S) was constant for any unknown samples analysed. The protease activity calculated as  $(U+S) - (U+B) - S$  showed identical trends to the protease activity calculated as  $(U+S) - (U+B)$ , therefore the latter formula was used in subsequent calculations to determine the true fluorescence increase due to protease activity. The protease activity was expressed in fluorescence change per unit sample as opposed to being calculated using a trypsin standard curve. The unknown samples originate from cell culture unclarified harvest; thus it is very likely that the samples contain one or more unknown proteases. In this case a standard curve might not be relevant. Plotting the true fluorescence increase due to protease activity versus the log of the dilutions used (Fig. 4-2 B)), the dilutions under which the colour effect was eliminated, were identified. It can be observed that between 1/10 to 1/100 dilution, there a relatively flat line, meaning that under these particular dilutions the colour effect is eliminated.





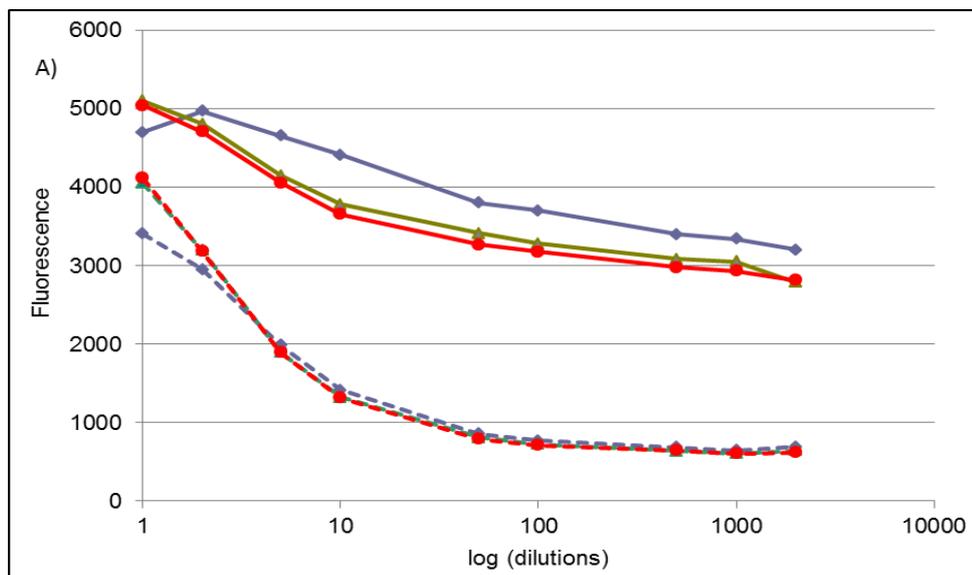
**Figure 4-2: Investigation into the colour effect of unknown samples. A) Fluorescence increase associated with the unknown sample containing the substrate (U+S) (—) and substrate-free control (U+B) (---), B) Fluorescence increase due to protease activity (U+S) – (U+B)**

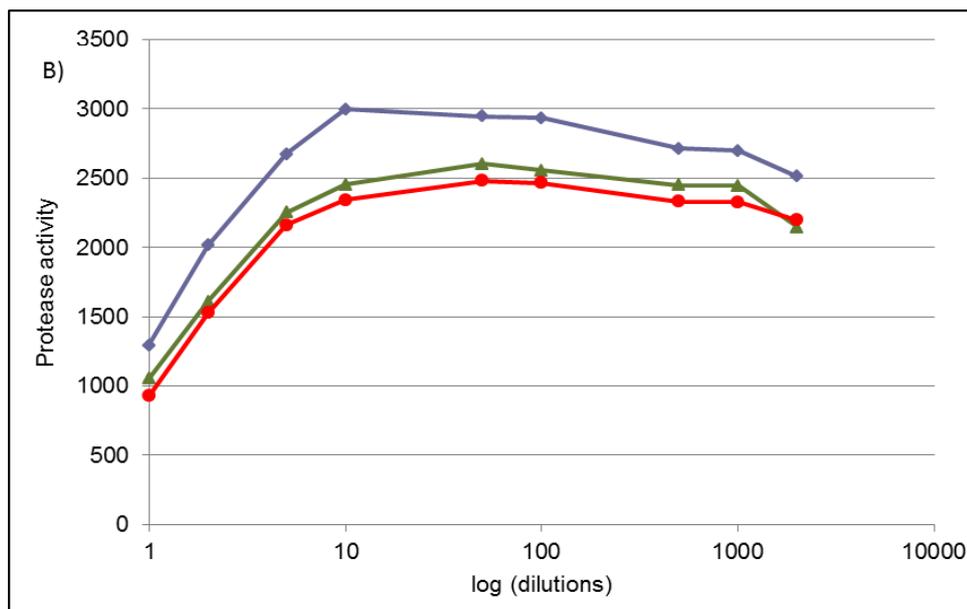
### 4.2.3 Effect of unclarified material on assay performance

Next, an investigation was carried out to determine whether the particles within unclarified harvest samples interfere in any way with the assay. A comparison was made between one unclarified and two clarified levels of the same unknown sample, at all the dilutions previously mentioned. Out of the two clarified samples, one was clarified only through centrifugation, whereas the other one was clarified through centrifugation followed by sterile filtration. The same as before, a fixed concentration of trypsin ( $3000 \text{ ng mL}^{-1}$ ) was incubated with casein substrate for 4 h in a 1:1 dilution until enzyme depletion. The concentration of trypsin used in combination with the substrate to reach enzyme depletion was the same as previously, in order to maintain consistency. This solution was then added to the different dilutions of the three levels of clarification and the corresponding fluorescence increase was read at time  $t = 0$  (U+S) (Fig. 4-3 A) —). For each dilution of each clarification level, a corresponding control sample without the casein substrate (U+B) was prepared as a 1:1 ratio of sample and digestion buffer (Fig. 4-3

A) ...). The true fluorescence increase due to protease activity was then calculated  $(U+S) - (U+B)$  and plotted against the log of dilutions (Fig. 4-3 B))

Similar trends are observed for all three samples, regardless of the levels of clarification. This suggests that the particles within unclarified samples do not interfere with the assay and the performance of the assay in determining protease activity is similar regardless of the level of clarification used. The fluorescence intensity is seen to be slightly higher for the unclarified sample. This is due to the higher initial colour intensity, for each dilution of this sample.





**Figure 4-3: Comparison between unclarified and clarified levels of the same unknown sample. Clarification was achieved through either centrifugation ( — ) or centrifugation followed by sterile filtration ( - - ). A) Fluorescence increase associated with the three levels of clarification of the unknown sample containing the substrate (U+S) ( — ) and substrate-free control (U+B) ( - - - ), B) Fluorescence increase due to protease activity (U+S) – (U+B), for each level of clarification**

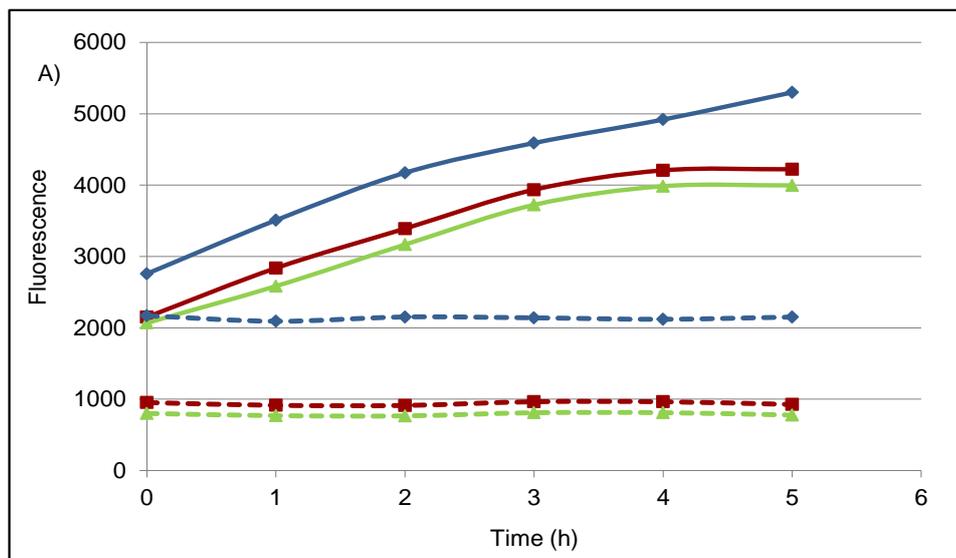
### 4.3 Analysis of unknown samples

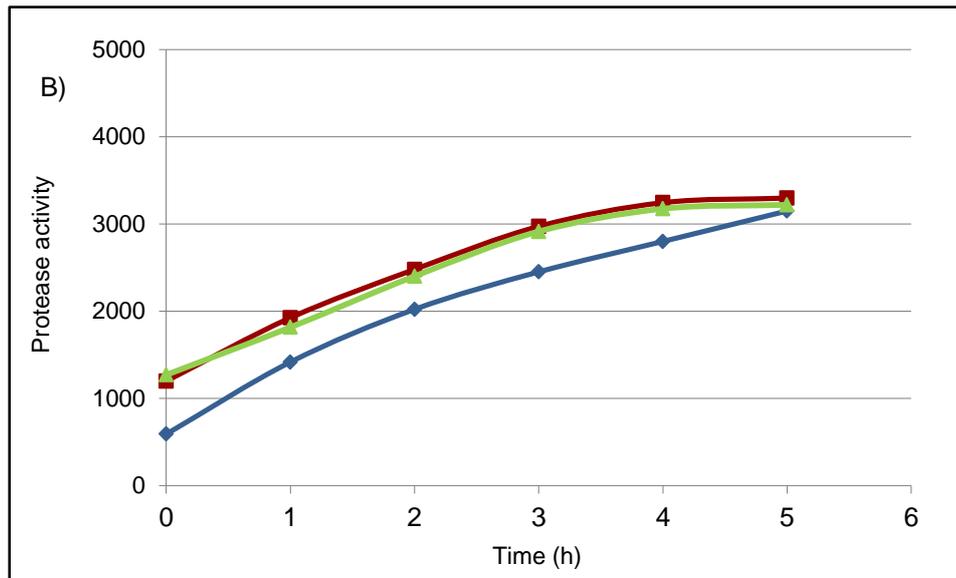
#### 4.3.1 Identification of suitable dilutions

Previously we identified 1/10, 1/50 and 1/100 dilutions as being the only dilutions from the ones investigated, under which the colour effect interference was not apparent. Next, we compared the protease activity associated with each dilution, for an unknown sample. Casein substrate has been added to each of the three dilutions of the unknown sample (U + S) and corresponding substrate-free control samples have also been prepared (U + B). The samples were added to a 96 well plate, kept in the dark at 40°C and read every hour for 5 h. The associated fluorescence increase of the samples with substrate (U+S) and of the substrate-free samples (U+B) for each dilution is shown in Fig. 4-4 A). The protease activity is then calculated by subtracting the substrate-free sample's fluorescence from the sample with substrate fluorescence and presented in Fig. 4-4 B). Fig. 4-4 shows that for the 1/50 and 1/100

dilutions, the trends are much more consistent than that for 1/10 dilution. For the purpose of obtaining consistent and reproducible results, 1/50 and 1/100 dilutions were chosen for subsequent analysis.

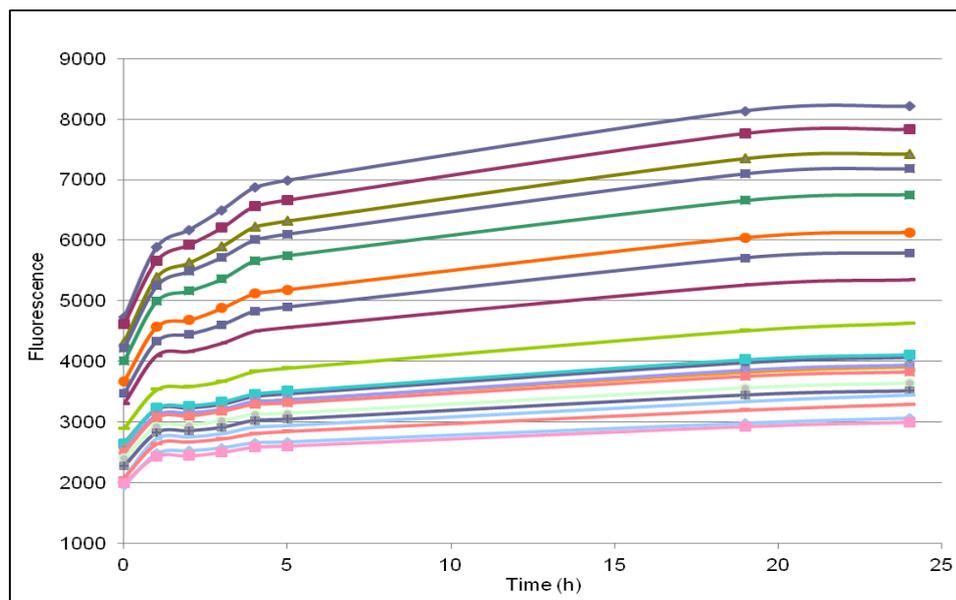
Two sets of aliquots from all AMBR 4 cultures were analysed for protease activity in separate days. The overall variability between the two analysis days was within 10% (data not shown). This did not have an impact on the trends observed, therefore the results of only one set of aliquots is presented in this chapter.





**Figure 4-4: Protease activity increase over time for an unknown sample. A) The fluorescence increase of the sample with substrate (U+S) (—) and substrate-free sample (U+B) (---) for dilutions 1/10 (—), 1/50 (—) and 1/100 (—). B) Fluorescence increase due to protease activity ((U+S) – (U+B)) for each of the three dilutions**

In order to determine the most suitable time for reading the plates, a trypsin standard curve was performed, with trypsin concentration varying from 50 – 10,000 ng mL<sup>-1</sup>. The plate was read every hour up to 5 h, then at 19 h and 24 h. We can see from Fig. 4-5 that the same distinction between the different levels of trypsin that can be seen all throughout 0-5 h is also maintained at the higher time points. For the purpose of analysing the plates in a timely manner, a time of 5 h was chosen for subsequent analysis using 1/50 dilution.



**Figure 4-5: Trypsin standard curve. The trypsin concentration varied from 50 – 10,000 ng mL<sup>-1</sup>, in-house Tris-HCl digestion buffer was used for making up the dilutions and the plate was incubated in the dark at 40° C. The increase in trypsin concentration is proportional to the increase in fluorescence**

### 4.3.2 Impact of cell culture inputs on protease activity

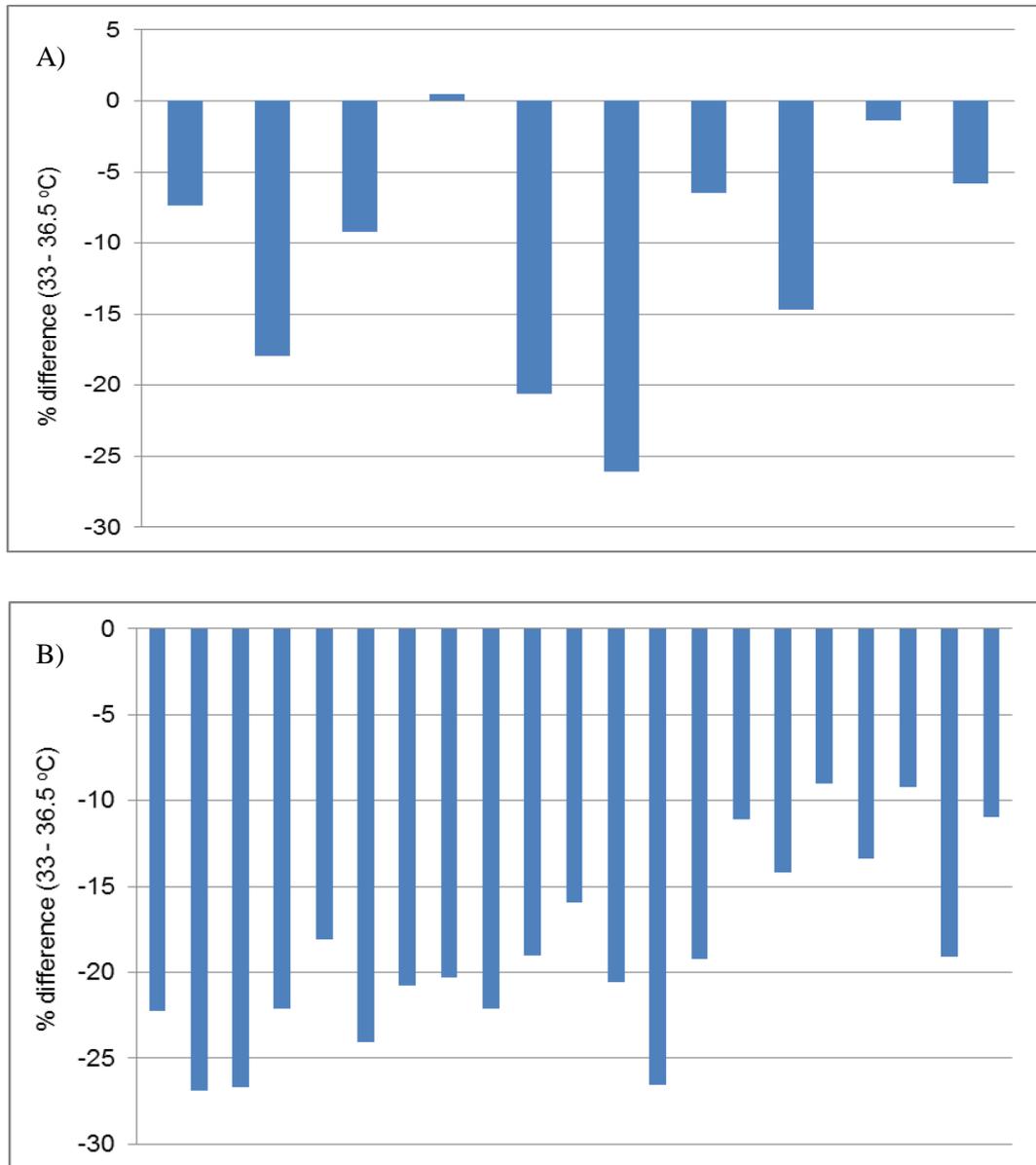
Published literature regarding proteases is mainly focused on the identification of these enzymes using mass spectrometry as part of downstream process development. This class of process-related impurities is of concern in the manufacturing process of therapeutic proteins. They are classified as problematic HCPs due to their ability to either associate with the product or due to their potential impact on the therapeutic protein's structural stability (Robert et al., 2009; Dorai et al., 2011; Dorai and Ganguly, 2014; Aboulaich et al., 2014). Only a few studies went a step further to actually quantify the amount of proteases present in CHO cultures (Robert et al., 2009; Dorai et al., 2011) but none of them investigated the relationship between protease activity and cell culture inputs (seeding density, media osmolality) and outputs (mAb titre, HCP levels) at harvest.

#### 4.3.2.1 Temperature

Lowering culture temperature from its optimum temperature for growth of 36.5 °C has been shown to improve recombinant protein productivity (Yoon et al., 2003; Tsao et al., 2005; Becerra et al., 2012) arrest cell growth, reduce lactate production, glucose consumption and ammonia production as well as to maintain higher culture viability for longer (Fox et al., 2004; Tait et al., 2013; Mason et al., 2014). Apart from having a beneficial impact on protein specific productivity, product quality has also been shown to be temperature dependent. Enhanced product quality (characterized by minimal impact on structural stability) at lower culture temperature has been associated with lower protease activity and the lower activity of various other temperature-dependent enzymes (Chuppa et al., 1997; Kaufmann et al., 1999). The effect of culture temperature on proteolytic activity present within bacterial and mammalian cell cultures has been widely studied in literature (Chuppa et al., 1997; Clark et al., 2004; Zhang and Chang, 2004; Dragomirescu et al., 2008). Chuppa et al. (1997) studied the effect of temperature (34, 35.5 and 37 °C) on proteolytic activity within high-density perfusion mammalian culture. They found that protease activity at 34 and 35.5 °C was similar, and lower than the protease activity at 37 °C. Zhang and Chang (2004) investigated the effect of temperature on proteolytic activity of human HtrA2 protease in *E.coli* culture over a wide range, 25 to 70 °C. The proteolytic activity of HtrA2 was seen to rapidly increase with temperature from 25 to 55 °C and then decreasing towards 70 °C. Qureshi et al. (2011) looked at optimising culture conditions for production of protease (for industrial purposes) by *Bacillus subtilis*. They tested a temperature range of 25 to 55 °C and found protease activity to increase up until 45 °C when it reached maximum and then decrease. In both cases, protease activity was found to be lower at a lower temperature.

Protease activity was measured at 1/50 dilutions within AMBR 3 (Fig. 4-6 A)) and AMBR 4 runs (Fig. 4-6 B)). The % difference in protease activity between replicate conditions at 33 °C compared to 36.5 °C was calculated. Within AMBR 3 cultures, at 33 °C, a 1.5 – 25 % decrease in protease activity was observed (Fig. 4-6 A)), while up to 26 % decrease in protease activity was seen in AMBR 4 cultures. This shows that regardless of the other combination of cell culture conditions present

within each culture, as part of the DoE design, culture temperature has a significant impact on proteolytic activity. A reduction in culture temperature could control the activity of these problematic HCPs and minimise the risk of product degradation due to proteolytic activity, hence maintaining product quality.

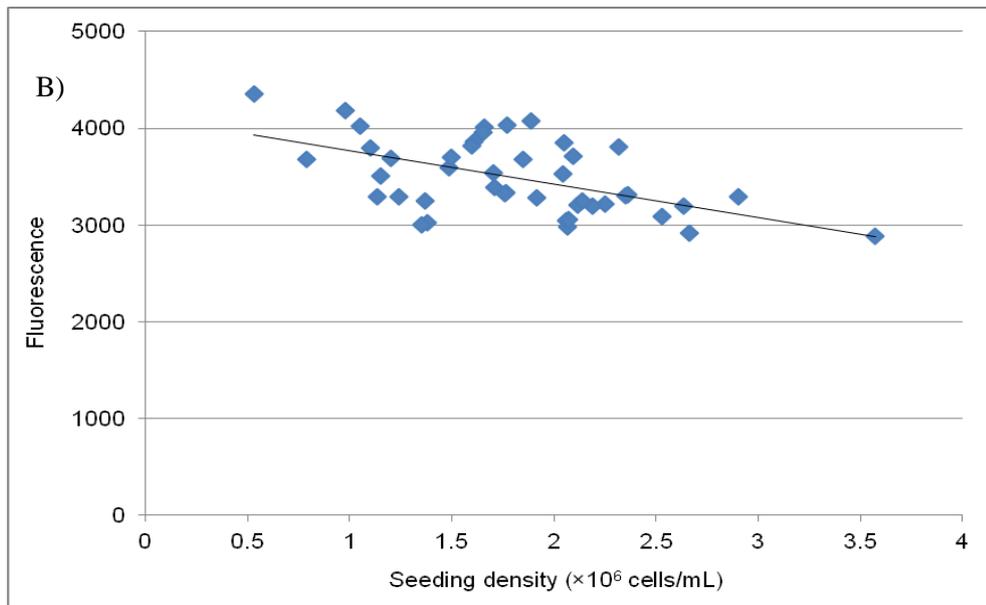
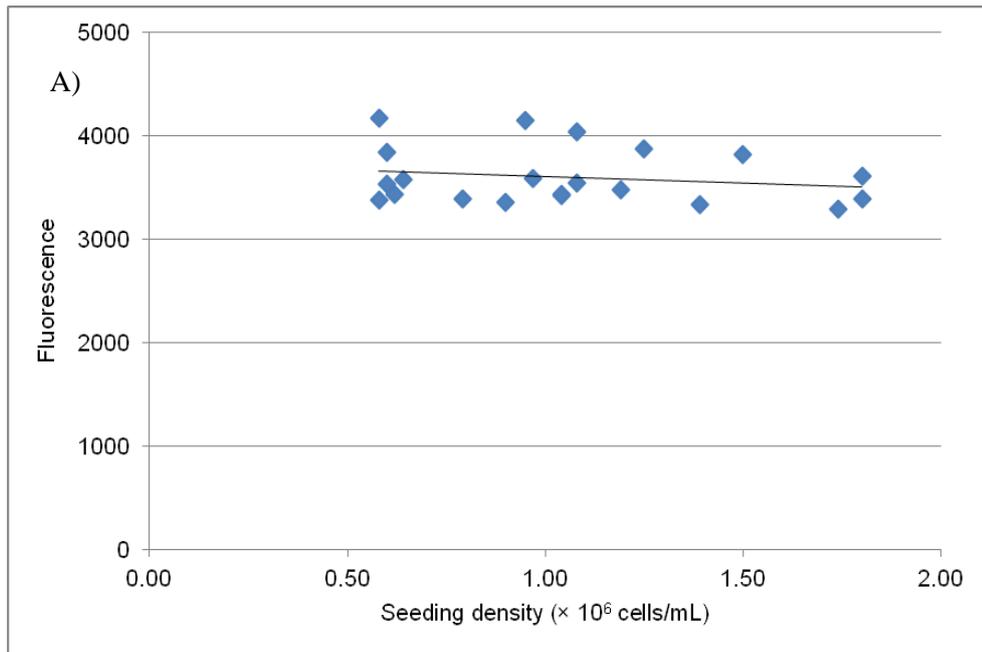


**Figure 4-6: Difference in protease activity (%) between equivalent conditions at 33 °C compared to 36.5 °C for (A) AMBR 3 and (B) AMBR 4**

#### 4.3.2.2 Seeding density

As part of the DoE designs for both AMBR 3 and AMBR 4, the influence of other cell culture inputs such as seeding density and starting media osmolality on protease activity could also be analysed. The influence of seeding density on protease activity has not been previously reported in literature. It can be observed from Fig. 4-7 A) that there is a constant level of protease activity (expressed as fluorescence increase) regardless of the seeding density used within AMBR 3 cultures. This suggests that for these cultures, the starting cell density within the DoE designs does not influence the protease activity in the harvest material.

In AMBR 4 cultures (Fig. 4-7 B)), a moderate negative correlation ( $R^2 = 0.41$ ) is observed between protease activity and seeding density. For AMBR 4 experiments, an ambr 48 (48 parallel cultures) was used as opposed to an ambr 24 (24 parallel cultures) for AMBR 3 experiments, therefore a wider seeding density range was able to be investigated. It is believed that the extended seeding density range ( $1.8 - 3.5 \times 10^6$  cells/mL) may be the cause of the negative trend observed between protease activity and seeding density within AMBR 4. Comparing the trend in AMBR 4 cultures that are under the same conditions as AMBR 3 (up to  $1.8 \times 10^6$  cells/mL seeding density), a similar constant protease activity was seen for the whole seeding density range, comparable to the trend seen in AMBR 3 (data not shown).

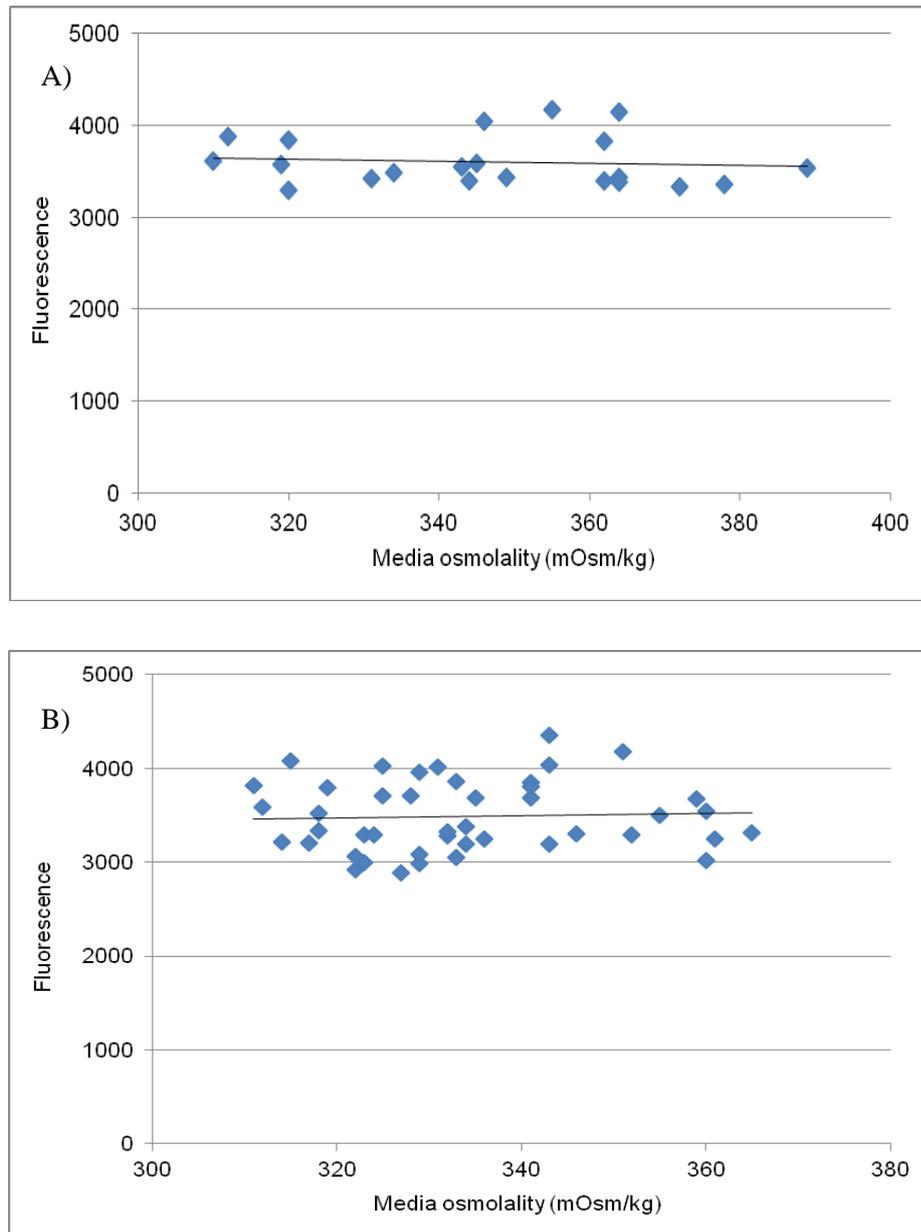


**Figure 4-7: The impact of seeding density on protease activity (measured as a fluorescence increase) within AMBR 3 (A) and AMBR 4 (B) cultures**

#### 4.3.2.3 Media osmolality

Media osmolality has generally been used in cell culture for its known ability to influence the specific cell productivity. Several published articles have shown that an elevated media osmolality can increase specific mAb productivity in CHO cultures. A high level of osmolality is also known to suppress cell growth (Ryu et al., 2000;

Kim et al., 2002; Lee et al., 2003; Li et al., 2010). The impact of media osmolality on protease activity was not previously investigated in literature. It seems that within the osmolality range explored, 310 – 390 mOsm/kg in AMBR 3 (Fig. 4-8 A)) and 310 – 370 mOsm/kg in AMBR 4 (Fig. 4-8 B)), protease activity is not affected by starting media osmolality.



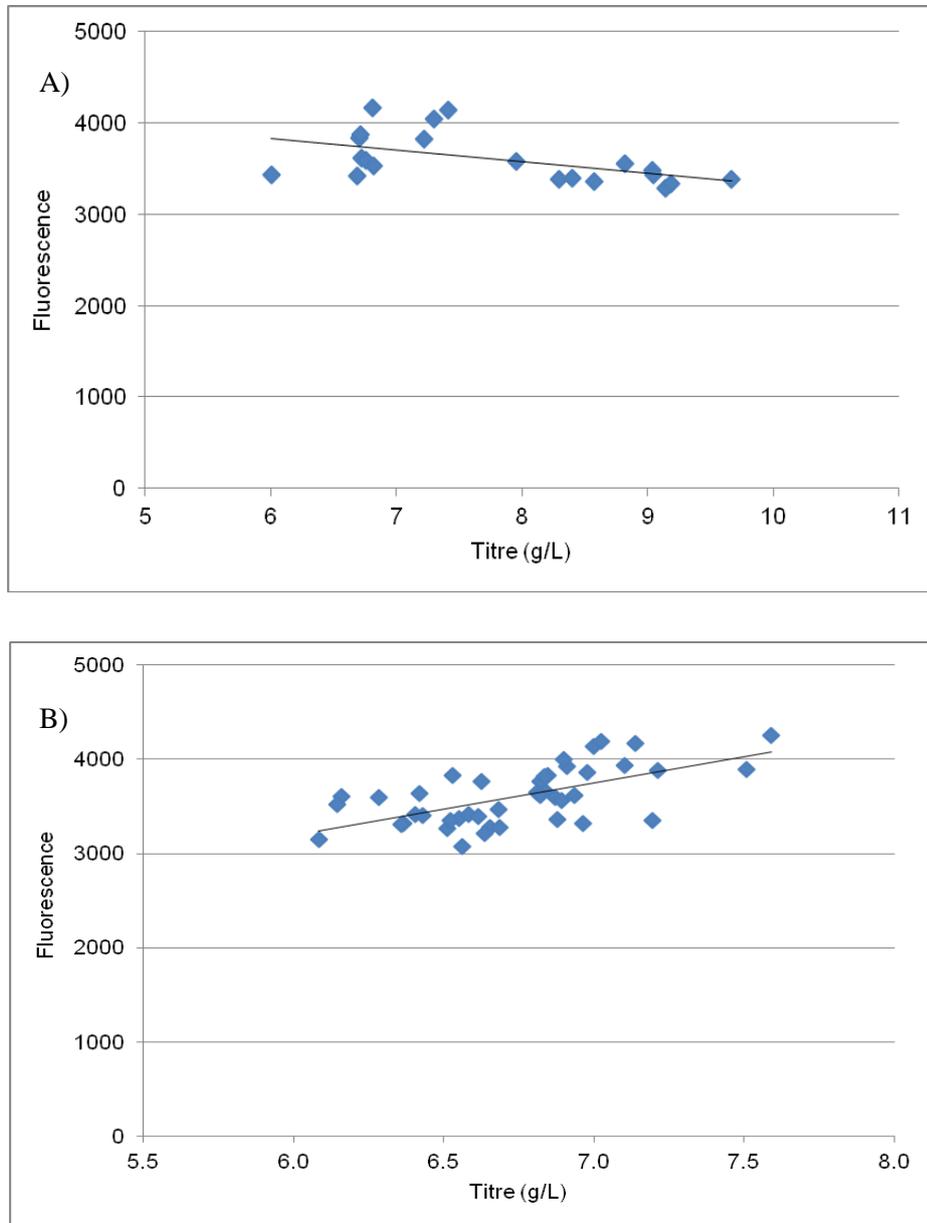
**Figure 4-8: The impact of starting media osmolality on protease activity (measured as a fluorescence increase) within AMBR 3 (A) and AMBR 4 (B) cultures**

### 4.3.3 Correlation between mAb titre, HCP levels and protease activity

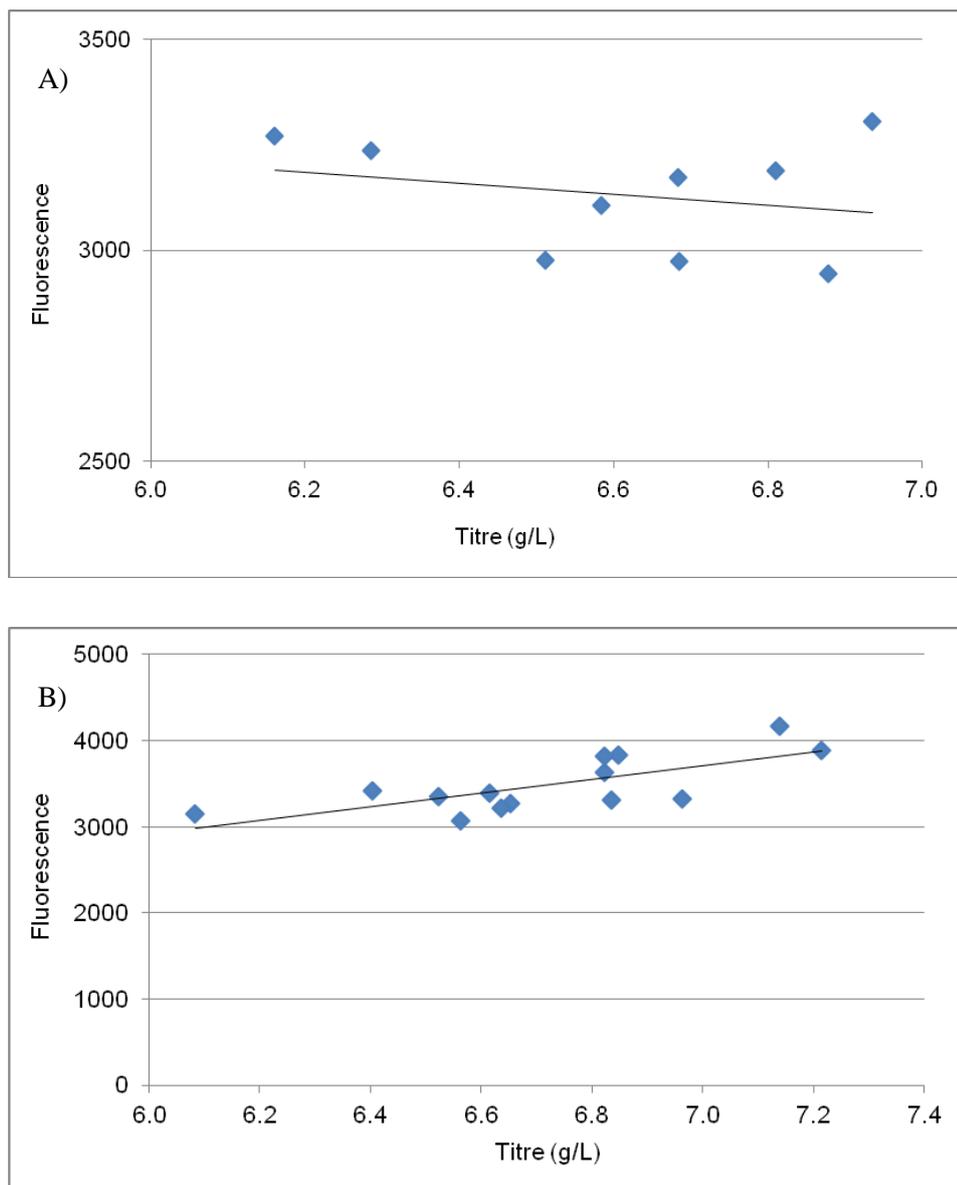
#### 4.3.3.1 mAb titre

The relationship between mAb titre and protease activity has not been previously examined in literature. Within AMBR 3 cultures, a weak negative correlation ( $R^2 = 0.28$ ) is seen between antibody titre and protease activity at harvest (Fig. 4-9 A)), while within AMBR 4 cultures, a moderate positive correlation is seen ( $R^2 = 0.45$ ) (Fig. 4-9 B)). In order to understand the difference in trends between the two AMBR runs, for AMBR 4, the protease activity was plotted against mAb titre for only the conditions that were part of the design space in AMBR 3 (pH 6.8 and up to  $1.8 \times 10^6$  cells/mL seeding density). A weak negative correlation was also seen between protease activity and titre in this case (Fig. 4-10 A)) which implied that the overall positive trend seen in AMBR 4 must be a consequence of the conditions not present within the AMBR 3 design space. The main difference between AMBR 3 and AMBR 4 design spaces is the addition of culture pH in AMBR 4. The relationship between protease activity and mAb titre within the additional design space of AMBR 4 was then assessed (Fig. 4-10 B)). A moderate positive correlation ( $R^2 = 0.54$ ) is seen which seems to be driving the overall AMBR 4 trend. There is limited literature on the impact of culture pH on CHO derived proteases. For CHO cultures, pH has been seen to have a significant impact on cell growth, mAb titre and HCP levels (Fig. 3-1 B); Fig. 3-2 AMBR 2) and due to its effect on these parameters it is likely that pH would also have a significant impact on protease activity at harvest. Culture pH has been known to have a significant impact on protease activity within bacterial cultures (Dragomirescu et al., 2008; Qureshi et al., 2011). Dragomirescu et al. (2008) investigated the pH dependence of protease activity produced in *Bacillus licheniformis* cells in the 3-12 range and seen an increase in protease activity up until pH 8, followed by a decrease. Qureshi et al. (2011) looked at optimising culture conditions for protease production by *Bacillus subtilis* and found that protease synthesis increased with an increase in initial culture pH and reached maximum at pH 8.5. For this cell line, within the design space investigated and at constant pH, a slightly negative correlation was identified between protease activity and mAb titre (Fig. 4-10 A)). Once culture pH was added to the experimental design, the significant

effect pH had on cell growth, mAb titre and HCP levels drove the positive relationship between protease activity and mAb titre within AMBR 4 cultures (Fig. 4-10 B)).



**Figure 4-9: The relationship between protease activity (measured as a fluorescence increase) and mAb titre within AMBR 3 (A) and AMBR 4 (B) cultures**



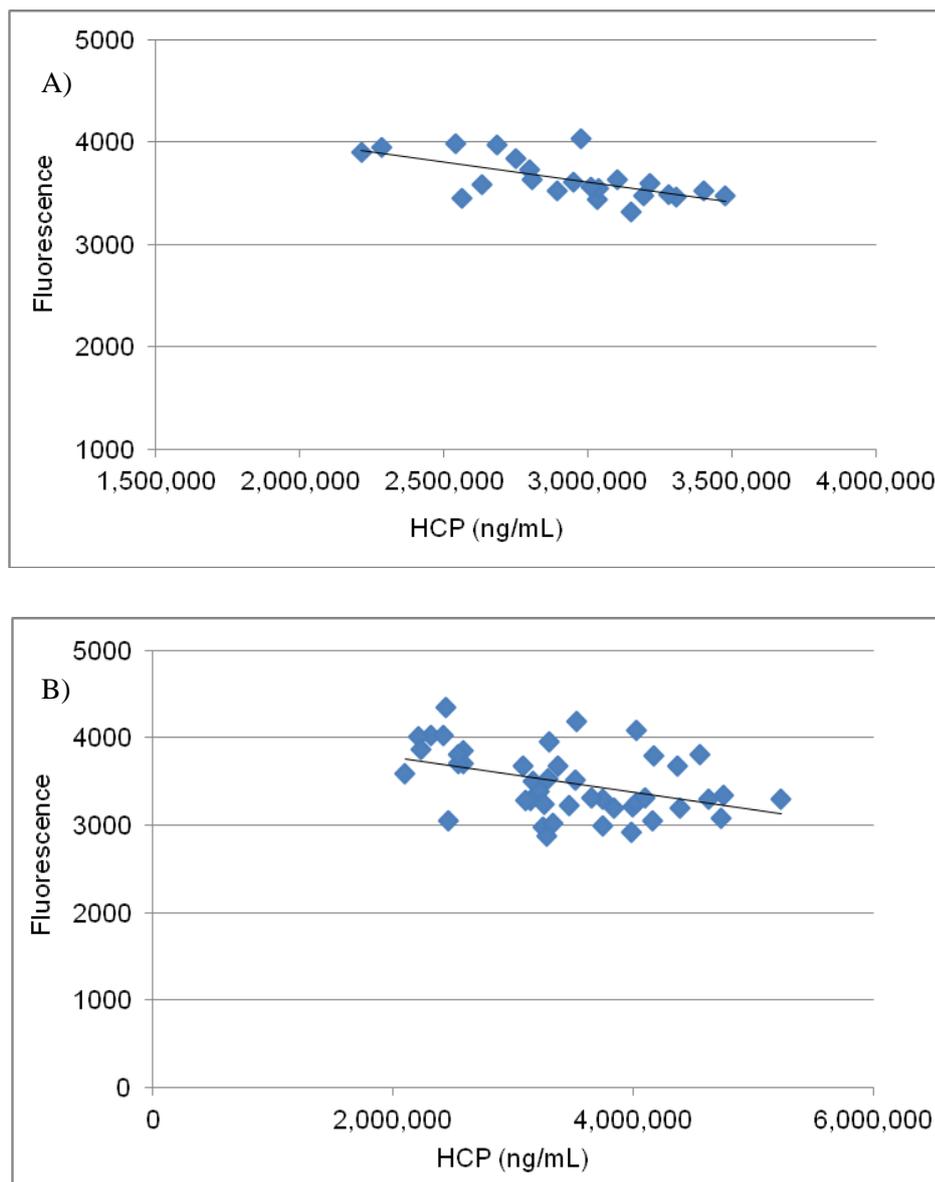
**Figure 4-10: The relationship between protease activity (measured as a fluorescence increase) and mAb titre within AMBR 4, split by pH. A) 6.8 B) 7-7.4**

#### 4.3.3.2 HCP levels

The abundance and composition of HCPs present in the harvest material can be influenced by a variety of factors, including cell culture and harvest conditions (Jin et al., 2010; Gutierrez et al., 2012; Tarrant et al., 2012). In Chapter 3 it was shown that different combinations of USP factors can have an impact on the resulting HCP levels. In recent years a lot of focus has been put on understanding the type of proteins that remain in the final drug product as well as the mechanism of how they

end up there. If we could understand the factors that influence the activity of specific problematic HCPs as well as identifying if their activity increases with an increase in overall HCP level we could potentially design processes that would ease the burden on the purification process. This would ensure a lower DSP process cost (DSP cost accounts for 45-92 % of the total cost of manufacturing a recombinant protein) resulting in a lower COG/g.

The hypothesis of protease activity increasing with an increase in total HCP levels was investigated. The relationship between the HCP and protease activity at harvest has not been previously looked at in literature. For both sets of experiments (AMBR 3 and AMBR 4) an increase in total HCP levels at harvest did not result in a concomitant increase in protease activity (Fig. 4-11 A), B)). As there are few published studies linking the levels of total HCP with levels of specific HCPs within the whole population, the levels of other types of problematic HCPs can increase with an increase in the overall HCP level. Yuk et al. (2015) investigated the relationship between phospholipase B-like-2 (PLBL2) and total HCP concentration in the supernatant and whole cell culture fluid, for three CHO null cell lines. PLBL2 is an HCP species present in CHO cultures, classified as problematic due to its ability to interact with certain recombinant humanized mAbs (Vanderlaan et al., 2015). They have seen an increase in PLBL2 concentration with an increase in HCP concentration, in all three null cell lines. It seems that different species within the whole HCP population are expressed differently in relation to the overall HCP trend. Efforts should be made to try and minimise the total level on HCP in harvest material, as this will reduce the risk of any immunogenic residual HCP to be found in the final product (Hogwood et al., 2014).



**Figure 4-11: The relationship between protease activity (measured as a fluorescence increase) and HCP levels within AMBR 3 (A) and AMBR 4 (B) cultures**

#### **4.4 Conclusions**

The impact of several cell culture parameters (temperature, seeding density, media osmolality) on protease activity was examined. It was shown that apart from temperature which has a negative impact on protease activity, the other USP conditions did not significantly affect protease activity. The relationship between protease activity and HCP levels was assessed and it was shown that an increase in HCP levels does not result in a similar increase in protease activity (considering the

proteases as a population). One of the protease assay's limitations is the fact that it considers the proteases as a whole population and is not able to distinguish between different classes of proteases. In order to pinpoint the specific proteases, more advanced proteomic studies such as mass spectrometry need to be used. Due to a lack of resources and time limitation, a more in-depth analysis of the composition of the protease population was not possible. It has not been yet evaluated how the levels of individual proteases might increase or decrease with regards to the HCP levels at harvest. This could help provide a better understanding of how USP factors influence the levels of potential problematic individual proteases that pass on to the DSP process.

The identification of cell culture inputs that have an impact on increasing problematic HCPs (protease) can help design processes that not only result in lower total HCP levels but also lower activity of specific, problematic HCPs. Determining the factors that have an impact on protease activity early in process development enables the possibility to improve the purification process to be able to reduce these species and potentially minimize the chance of residual protease that could cause product damage or eliciting an immune response in patients. Manipulating the USP conditions in order to reduce the activity of problematic HCPs would have a benefit in reducing the cost of goods by minimising the burden on DSP or through the ability to implement alternative, more cost effective methods.

## **Chapter 5**

### **5 Derivation of predictive cause-and-effect correlations for cell culture using multivariate data analysis**

#### **5.1 Introduction**

The optimisation of cell culture processes in order to maximize antibody production as well as to reduce the time to market is an important aim within the biopharmaceutical industry. An important task in defining an optimisation strategy is to identify and predict cell culture behaviour (De Alwis et al., 2007). Cell culture experimentation involving trial and error optimisation of cell culture parameters, results in a large number of experiments, which can be time-consuming and expensive (Ho et al., 2006). Process modelling helps identify and evaluate product and process variables that might be critical to product quality and performance, facilitating a better planning and design of experiments. Modelling enables knowledge-based decision making, resulting in increased development efficiencies, continuous quality improvements and cost reduction (O’Kelly et al., 2012).

The aim of this chapter was to characterise the high throughput cell culture data generate using the ambr system (Chapter 3) by multivariate data analysis techniques (multiple linear regression, all possible stepwise regression) to derive statistical cause-and-effect correlations. These statistical equations are able to predict cell culture outputs such as mAb titre and HCP levels at harvest, based on cell culture inputs (temperature, media osmolality, seeding density, pH and timing of feed initiation) (Fig. 5-1). ANOVA analysis, using Design Expert and JMP software, was used to determine the significance of each parameter, either individually or in interaction with others on titre and HCP levels.

#### **5.2 Results and discussions**

For identifying the most suited predictive models for antibody titre and host cell proteins (HCP) from each ambr run, a stepwise all possible regression analysis (as

described in section 2.5.2.4) was used to narrow it down to eight hierarchical, best models, with increasing number of model terms. Stepwise all possible regression analysis tests all possible subsets of the set of potential independent variables. An analysis of variance is performed for each model to identify the main factors with a significant influence on the corresponding response (antibody titre or HCP). The titre and HCP results were analysed using JMP and Design Expert. The Analysis of Variance (ANOVA) is presented in Fig. 5-2 for titre and Fig. 5-5 for HCP, for all three ambr runs. The significant factors and/or interactions are presented with a \* (Table 5-1, 5-4). The selection of an optimal model should be a balance between maximising predicted  $R^2$  (described in section 2.5.6.2),  $R^2$  (described in section 2.5.6.1) as well as minimising RMSE (described in section 2.5.6.3) and AIC (described in section 2.5.3). Regression, like most statistical techniques, has a set of underlying assumptions that are expected to be in place if we want the estimated model to be reliable. For the selected best model for each response, the underlying assumptions are verified. Within all the subsequent analysis, the variables media osmolality, seeding density, temperature, pH and feed start time are referred to in coded factors as A, B, C, D and E respectively. This is for an easier interpretation of the results. The optimal models for antibody titre and HCP from AMBR 1 (low generation) are presented in more detail below, but a similar analysis was performed for titre and HCP from AMBR 2 and AMBR 3.

## **5.2.1 Titre prediction models from AMBR runs**

### **5.2.1.1 Model building and selection**

The best eight models resulting from the stepwise all possible regression analysis, for estimating antibody titre from AMBR 1 run are presented in Table 5-1 with their corresponding statistics. The R-squared term ( $R^2$ ) defined in equation (9) (Section 2.5.6.1) is an indication of how well the model fits the experimental data, while the adjusted  $R^2$  also takes into account the number of factors evaluated. This is helpful when comparing amongst models that were developed from datasets with different number of factors. The predicted  $R^2$  is calculated using regression analysis and is used to indicate how well the model can predict responses for new combinations of

variables within the same range evaluated to develop the model. The  $R^2$  k fold (Section 2.5.5.1) is a measurement of internal validation of the model and is an indication of how accurate and valid the statistical model is at predicting new data, within the same design space but different than the data used to initially build the model. RMSE refers to the model residuals while AIC is a model selection criteria used to compare between potential models. The focus for the resulting models is towards prediction; therefore predicted  $R^2$  and  $R^2$  k fold are critical statistics that have to be taken into account when selecting the best model. Their value should be maximised for a model with the best prediction capability.

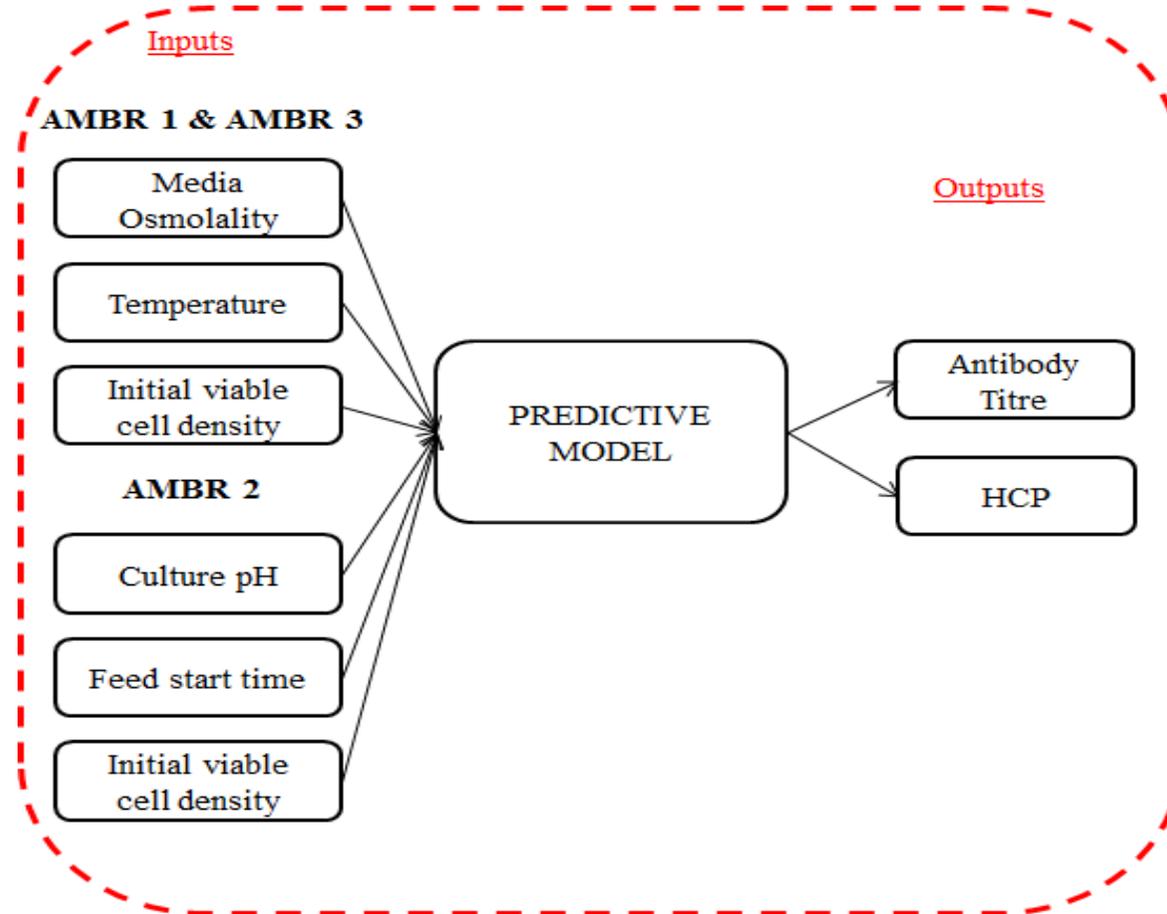


Figure 5-1: Integration of AMBR run experiments with predictive modelling

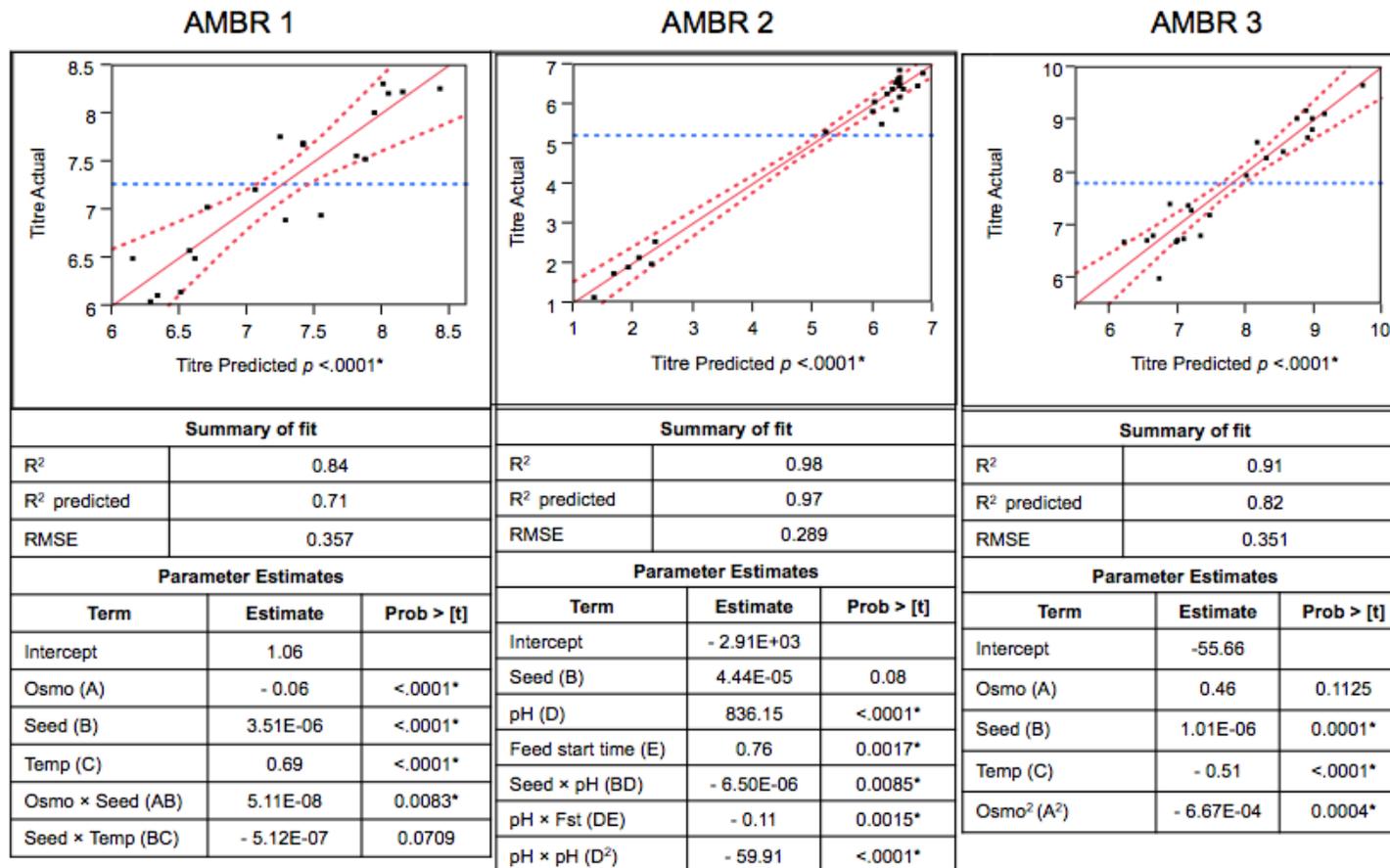


Figure 5-2: Analysis of mAb titre models for AMBR 1, 2 and 3. The analysis of variance as well as the correlation between experimental values and predictions from each model is presented \* Significant in terms of probability (p-value < 0.05)

**Table 5-1: Summary statistics for eight “best” models for AMBR 1 titre presented in coded factors. The coded factors A, B and C refer to the cell culture variables media osmolality, seeding density and temperature, respectively**

	AMBR 1 – Titre	STATISTICS					
Model terms	Model	R <sup>2</sup>	Predicted R <sup>2</sup>	Adjusted R <sup>2</sup>	R <sup>2</sup> k fold	RMSE	AIC
1	B*	0.25	0.07	0.21	0.07	0.685	47.01
2	B*+C*	0.48	0.27	0.41	0.29	0.591	43.12
3	A*+B*+C*	0.71	0.49	0.65	0.56	0.454	34.96
4	A*+B*+C*+AB*	0.80	0.60	0.74	0.64	0.390	31.79
<b>5</b>	<b>A*+B*+C*+AB*+BC</b>	<b>0.84</b>	<b>0.71</b>	<b>0.79</b>	<b>0.72</b>	<b>0.357</b>	<b>31.78</b>
6	A*+B*+C*+AB*+BC*+B <sup>2</sup>	0.85	0.72	0.79	0.71	0.355	35.85
7	A*+B*+C*+AB*+AC+BC+B <sup>2</sup>	0.86	0.58	0.78	0.55	0.359	41.56
8	A*+B*+C*+AB*+AC+BC+A <sup>2</sup> +B <sup>2</sup>	0.87	0.49	0.77	0.38	0.372	49.65

Out of the eight models shown in Table 5-1, model 6 has the highest predicted  $R^2$  of 0.72, closely followed by model 5 with a predicted  $R^2$  of 0.71. The  $R^2$  for model 6 is also slightly higher than model 5, with 0.85 compared to 0.84. The addition of more and more variables to a model will always increase the  $R^2$ , but when the increase in  $R^2$  is very small it is not worth increasing the complexity of the model. Model 5 is considered to be the best model as the difference between model 5 and 6 is very small in terms of  $R^2$  and predicted  $R^2$ . Model 5 also has a slightly higher  $R^2$  k fold of 0.72 compared to 0.71 for model 6. The selected model for antibody titre from AMBR 1 is analysed in more detail below. A similar analysis was performed for antibody titre from AMBR 2 (APPENDIX Table 5-1A) and AMBR 3 (APPENDIX Table 5-3A). The analysis of variance of model 5, the reduced 2FI (two-factor interaction) model, indicates that the significant model terms are media osmolality (A), seeding density (B), temperature (C), the interaction between seeding density and media osmolality (AB) and to a lower extent the interaction between media osmolality and temperature (BC). The  $p$ -value of the model of  $<0.0001$  implies that the model is significant (Fig. 5-2). The actual vs predicted graph is presented in Fig. 5-2. The ANOVA analysis for the best model for titre from AMBR 2 and AMBR 3 as well as their corresponding actual vs predicted graphs are also presented in Fig.5-2. The final equations for AMBR 1 titre in coded and actual factors are presented below:

$$\text{Titre} = \beta_0 + \beta_1 A^* + \beta_2 B^* + \beta_3 C^* + \beta_4 AB^* + \beta_5 BC \quad (17)$$

$$\begin{aligned} \text{Titre} = & 1.06 - 0.06 \times \text{Osmolality} + 3.51 \times 10^{-6} \times \text{Seeding density} + 0.69 \times \\ & \text{Temperature} + 5.11 \times 10^{-8} \times \text{Osmolality} \times \text{Seeding density} - 5.12 \times 10^{-7} \times \\ & \text{Seeding density} \times \text{Temperature} \end{aligned} \quad (18)$$

The model's coefficients are an indication of how different cell culture parameters influence mAb titre. This is useful for USP scientists to easily identify which parameters have the biggest impact and need close monitoring for maximising mAb titre. Media osmolality is seen to have a significant negative impact on titre, an increase in starting media osmolality results in lower titres. On the other hand, seeding density and temperature have a positive impact on titre, an increase in both

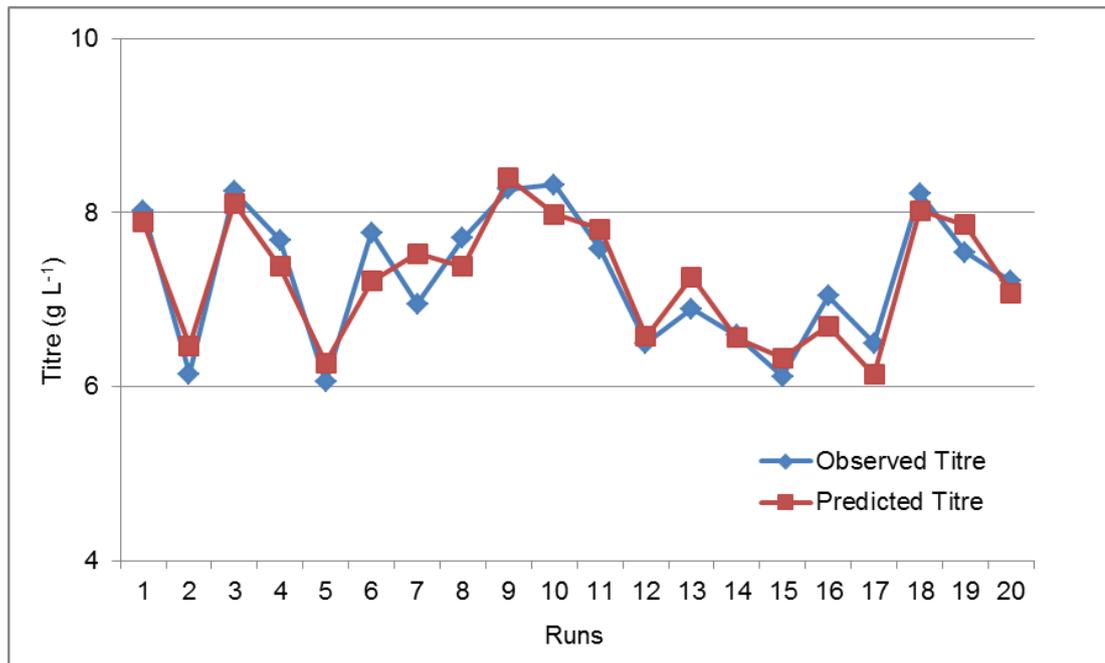
parameters leads to an increase in the overall titre. The highest titre within AMBR 1 is achieved for the lowest media osmolality and the highest seeding density and temperature within the design space explored. Another significant term is the interaction between media osmolality and seeding density. The negative effect of an increase in media osmolality on its own, starts to decline with an increase in seeding density. In terms of titre, cultures seeded at a high density are less sensitive to changes in osmolality than cultures seeded at a low density (Fig. 3-2).

In order to identify if the variability within the Protein A assay (Chapter 2, Section 2.4.4) has an impact on the signal to noise ratio within the AMBR 1 titre model, the model's coefficient of variation (% CV) was calculated and then compared to the CV of the assay. The difference between the assay variability (5 %) and the calculated % CV of the AMBR 1 titre model (4.92 %) was insignificant which implies that the predictive capability of the model was not influenced by the variability within the Protein A assay. An insignificant difference was also seen between the variability of titre models for AMBR 2 and AMBR 3 and the assay's inherent variation.

Abu-Absi et al. (2010) and Rouiller et al. (2012) both used QbD principles to characterise cell culture processes, by adopting a risk assessment exercise to identify potential critical and key process parameters with possible impact on process performance. They then used a DoE approach to evaluate the identified parameters and generate regression models for product titre (mAb and Fc-Fusion protein, respectively) as well as certain product and process related impurities. The comparison between the titre models generated in this work and the models they present are shown in Table 5-2. The comparison looks at the main parameters having a significant impact on product titre as well as  $R^2$  and  $R^2$  predicted. The models discussed here are comparable to the models presented in literature in terms of how well the models are able to fit the experimental data and how well they can predict for new combination of factors within the same design space. The models include process parameters not found within the literature models, such as media osmolality and feed start time.

**Table 5-2: Comparison between generated titre models and titre models found in literature. The comparison looks at the main parameters having a significant impact on product titre as well as  $R^2$  and  $R^2$  predicted**

TITRE			
Literature comparison	Significant factors	$R^2$	$R^2$ predicted
AMBR 1 (Low generation)	<ul style="list-style-type: none"> <li>• Media osmolality</li> <li>• Seeding density</li> <li>• Temperature</li> </ul>	0.84	0.71
AMBR 2 (Mid generation)	<ul style="list-style-type: none"> <li>• Seeding density</li> <li>• pH</li> <li>• Feed start time</li> </ul>	0.98	0.97
AMBR 3 (High generation)	<ul style="list-style-type: none"> <li>• Seeding density</li> <li>• Temperature</li> </ul>	0.91	0.82
Abu-Absi et al. (2010) Product: mAb	<ul style="list-style-type: none"> <li>• Seed bioreactor temperature</li> <li>• Production bioreactor T shift timing</li> <li>• Production bioreactor DO</li> <li>• Production bioreactor initial VCD</li> </ul>	0.88	0.61
Rouiller et al. (2012) Product: Fc-Fusion protein	<ul style="list-style-type: none"> <li>• pH</li> <li>• DO</li> <li>• Culture duration</li> <li>• Seeding density</li> </ul>	0.96	0.94



**Figure 5-3: Comparison between observed and predicted values of AMBR 1 titre. Runs 1-11 correspond to CS1 within AMBR 1, maintained at a temperature of 36.5 °C and runs 11-20 correspond to CS2 within AMBR 1, maintained at a temperature of 36.5 °C followed by a temperature shift to 33 °C in day 4 of culture**

Once a model has been chosen, it needs to be tested in order to ensure that the model is capable of predicting new data, within the same design space. This can be done through internal and external validation. Internal validation refers to the ability of the chosen model to predict data that has already been part of the experimental design but was not used to construct the model (k-fold cross validation). The  $R^2$  k fold term has been already used as a selection criterion when choosing the “best” model. Using the chosen model, titre values have been predicted for the experimental design already performed. The observed vs predicted graph (Fig. 5-3) shows that the model chosen gives a good prediction of experimental values. Ideally, an external validation of the model would also be performed. An external validation involves using the model to predict titre values associated with combinations of input variable (within the same design space) that have not been previously performed as part of the experimental design. New sets of experiments would be run and the final measured titre values would be compared to the already predicted ones in order to assess the

model's ability of externally predicting titre. As further experimentation was not possible within the time provided, only internal validation was used to assess the validity of the models (for both titre and HCP).

### 5.2.1.2 Verifying assumptions

In order to test the normality of residuals the Shapiro-Wilk test (as described in section 2.5.4.1) was performed. The Shapiro-Wilk statistics for titre and its corresponding  $p$ -value are presented in Table 5-3. The  $p$ -value of 0.61 is not statistically significant ( $p$ -values values lower than 0.05 are generally considered statistically significant) therefore the hypothesis of normality of residuals should not be rejected. This suggests that the assumption is not violated and the residuals are normally distributed.

**Table 5-3: The statistics for the Shapiro – Wilk and Durbin – Watson Tests and their corresponding  $p$ -value for AMBR 1 Titre and HCP**

	Shapiro- Wilk W Test		Durbin – Watson Test	
	W	$p$ -value*	Durbin - Watson	$p$ -value*
Titre	0.96	0.61	2.92	0.94
HCP	0.91	0.06	2.24	0.57

\* A  $p$ -value  $< 0.05$  is generally considered statistically significant

To verify whether or not the residuals are correlated the Durbin-Watson autocorrelation test (as described in section 2.5.4.2) was performed. The Durbin – Watson statistics for titre and its corresponding  $p$ -value are presented in Table 5.3. The  $p$ -value of 0.94 is not included within the two-tailed rejection region of  $p$ -value  $\leq 0.025$  or  $p$ -value  $\geq 0.975$ , therefore the null hypothesis of no autocorrelation is not rejected. The residuals are not correlated.

To check the linearity of the model, the residual vs predicted plot is evaluated (Fig. 5-4). The residuals are evenly distributed on either side of the regression line and there is no visible trend. The assumption of linearity is satisfied.

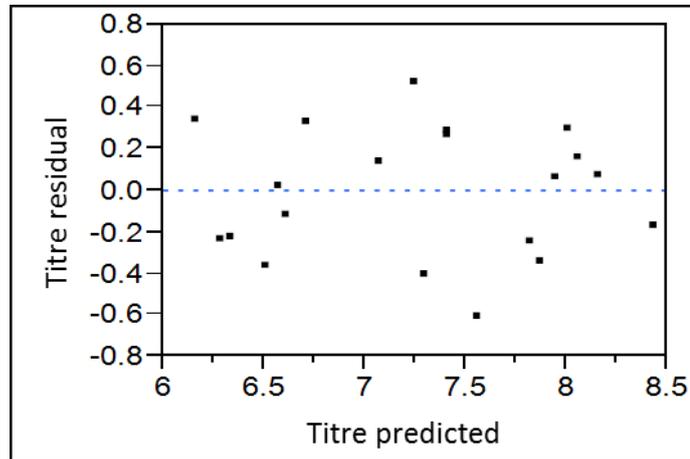


Figure 5-4: Residual versus predicted values AMBR 1 titre

**Table 5-4: Summary statistics for eight “best” models for AMBR 1 HCP presented in coded factors. The coded factors A, B and C refer to the cell culture variables media osmolality, seeding density and temperature, respectively**

	AMBR 1 – HCP	STATISTICS					
Model terms	Model	R <sup>2</sup>	Predicted R <sup>2</sup>	Adjusted R <sup>2</sup>	R <sup>2</sup> k fold	RMSE	AIC
1	C*	0.63	0.55	0.61	0.57	214836	553.26
2	B+C*	0.66	0.51	0.62	0.52	213495	555.03
3	A+C*+A <sup>2</sup>	0.70	0.54	0.65	0.57	205486	555.91
4	<b>A+B+C*+A<sup>2</sup></b>	<b>0.72</b>	<b>0.50</b>	<b>0.65</b>	<b>0.51</b>	<b>205141</b>	<b>558.72</b>
5	A+B+C*+AB+A <sup>2</sup>	0.78	0.48	0.70	0.38	189332	559.01
6	A+B+C*+AB+BC+A <sup>2</sup>	0.78	0.40	0.68	0.28	194900	564.44
7	A+B+C*+AB+AC+BC+A <sup>2</sup>	0.78	0.18	0.66	-0.02	202770	571.33
8	A+B+C*+AB+AC+BC+A <sup>2</sup> +B <sup>2</sup>	0.78	-0.19	0.62	-0.36	211784	579.78

\* Significant value in terms of probability ( $p$ -value < 0.05)

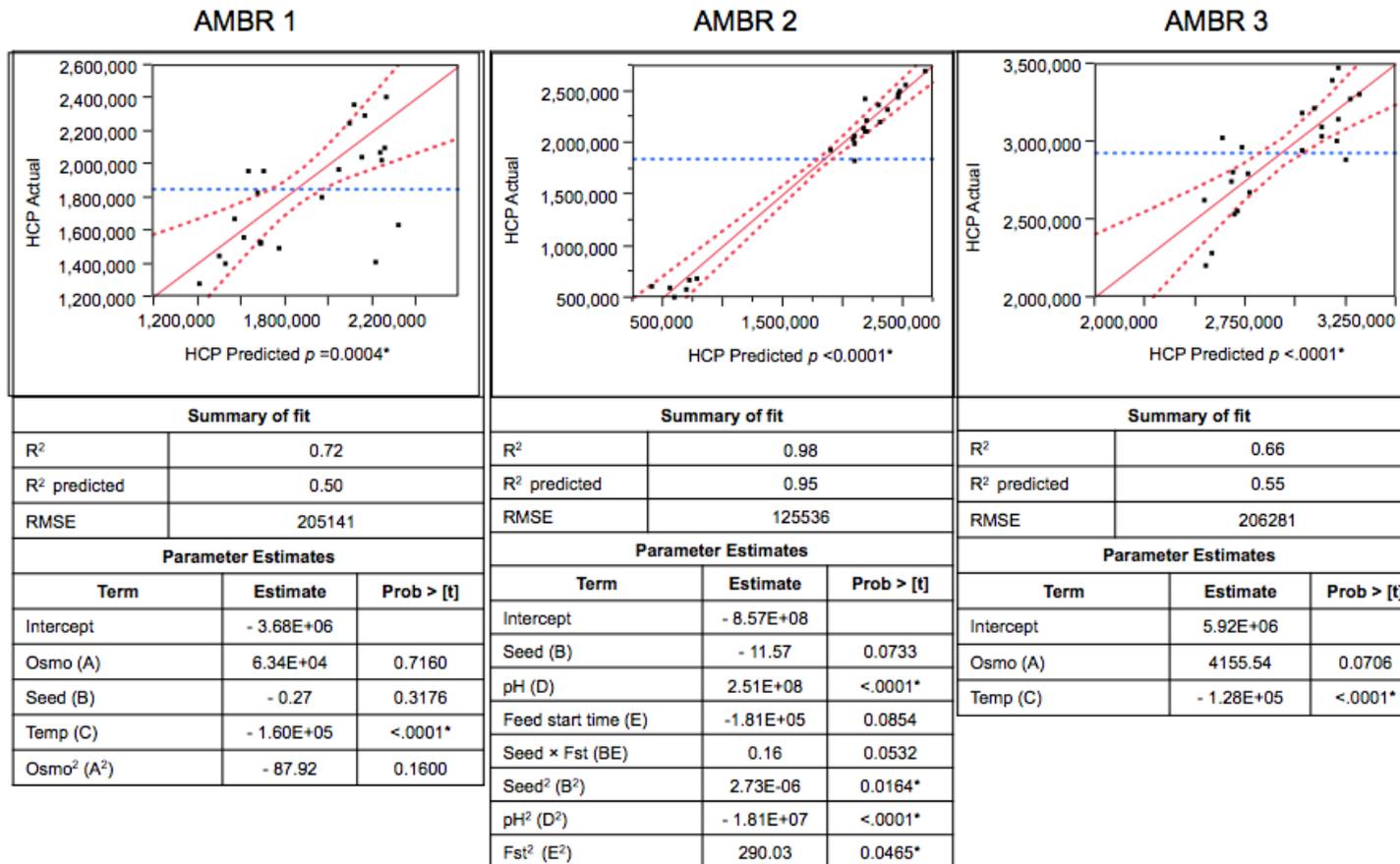


Figure 5-5: Analysis of HCP models for AMBR 1, 2 and 3. The analysis of variance as well as the correlation between experimental values and predictions from each model is presented \* Significant in terms of probability ( $p$ -value < 0.05)

## 5.2.2 HCP prediction models from AMBR runs

### 5.2.2.1 Model building and selection

The best eight models for estimating HCP from AMBR 1 run, with their corresponding statistics are presented in Table 5-4. Out of the eight models resulting from the all possible stepwise regression, model 1 has the highest predicted  $R^2$  of 0.55 closely followed by model 3 with a predicted  $R^2$  of 0.54 and models 2 and 4 with a predicted  $R^2$  of 0.51 and 0.50, respectively. Although model 1 has the highest predicted  $R^2$  it also has a poor fit, shown by its low  $R^2$  value of 0.63. This is due to the fact that model 1 has only one predictor for HCP (temperature) not taking into account any other factors within the design space that might affect HCP levels. This suggests that model 1 might not be the most suited for HCP prediction. The second model in terms of prediction capability is model 3, with a predicted  $R^2$  of 0.54 and a  $R^2$  of 0.70. Upon testing the underlying assumptions of this model, the normality of residuals assumption was violated, making it an unsuited model for the prediction of HCP levels. Out of the two models that follow closely in terms of prediction, model 2 and 4, model 4 fits the data better with a  $R^2$  of 0.72 compared to a  $R^2$  of 0.66 for model 2, as well as having a lower error, reason for which it is considered to be the best model.

The selected model for HCP from AMBR 1 is analysed in more detail below. A similar analysis was performed for HCP from AMBR 2 (APPENDIX Table 5-2A) and AMBR 3 (APPENDIX Table 5-4A). The analysis of variance of model 4 indicates that the significant model term is temperature (C), with media osmolality (A), seeding density (B) and the quadratic term for media osmolality ( $A^2$ ) having a much lower contribution to the overall model. The model's  $p$ -value of 0.0004 implies that the model is significant. The actual vs predicted graph is shown in Fig.5-5 for AMBR 1. ANOVA analysis for the best HCP model for AMBR 2 and AMBR 3 as well as their corresponding actual vs predicted graphs are also presented in Fig.5-5.

The final equations for AMBR 1 HCP in coded and actual factors are presented below:

$$\text{HCP} = \beta_0 + \beta_1 A + \beta_2 B + \beta_3 C^* + \beta_4 A^2 \quad (19)$$

$$\text{HCP} = - 3.68 \times 10^6 + 6.34 \times 10^4 \times \text{Osmolality} - 0.27 \times \text{Seeding density} - 1.60 \times 10^5 \times \text{Temperature} - 87.92 \times \text{Osmolality}^2 \quad (20)$$

The model's coefficients are an indication of how different cell culture parameters influence HCP levels at harvest. The main significant factor influencing HCP levels is culture temperature (C), presenting the higher coefficient within the regression equation. Culture temperature is seen to have a significant negative impact on HCP levels, an increase in culture temperature leading to a lower HCP levels. Media osmolality and seeding density have a much lower effect on HCP levels, with osmolality having a small positive effect and seeding density a small negative effect. The lowest HCP level within AMBR 1 is achieved for high culture temperature (36.5 °C), base osmolality (314 mOsm kg<sup>-1</sup>) and high seeding density (1.14 × 10<sup>6</sup> cell mL<sup>-1</sup>). The same set of conditions also correspond to the highest mAb titre achieved within AMBR 1, which indicates that within this set of experiments, no compromise has to be made between high mAb titre and low HCP levels.

In order to identify if the variability within the Gyros HCP assay (Chapter 2, Section 2.4.5) has an impact on the signal to noise ratio within the AMBR 1 HCP model, the model's coefficient of variation (% CV) is calculated and then compared to the CV of the assay. The difference between the assay variability (20 %) and the calculated % CV of the AMBR 1 HCP model (7 %) is significant which implies that the model's ability to predict new data is affected by the high HCP assay variability. This is also reflected in the lower R<sup>2</sup> predicted of the HCP models compared to the titre models.

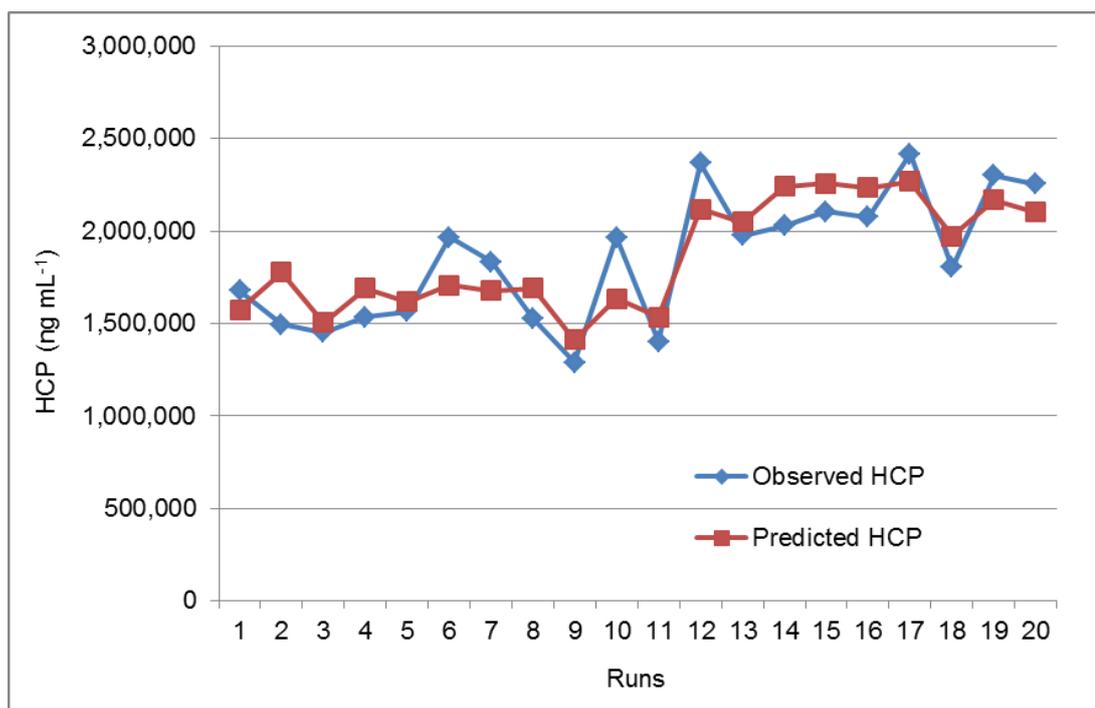
Rouiller et al. (2012) presented a HCP regression model for the production bioreactor of CHO cells expressing an Fc-Fusion protein. The most significant factor affecting HCP levels was pH and to a lesser extent DO and culture duration. The

comparison between this model and the HCP models generated from AMBR 1, 2 and 3 data is presented in Table 5-5. The model presented by Rouiller et al. (2012) had an  $R^2$  of 0.92 and an  $R^2$  predicted of 0.88, which is higher than the statistics for the AMBR 1 and 3 models, but more similar to the AMBR 2 HCP model. The HCP models presented here include process parameters not previously used in literature to predict HCP, such as media osmolality, seeding density, culture temperature and feed start time. The analysis of variance for mAb titre and HCP levels, within AMBR 1 experiments, indicates that the main factor with a significant influence on both titre and HCP is culture temperature. Culture temperature has a significant positive effect on mAb titre and a negative effect on HCP levels. At a high culture temperature (36.5 °C), a high titre and low HCP level is achieved within AMBR 1.

The final predictive models for titre and HCP from each AMBR run are presented in Table 5-6 in coded factors. The HCP models presented here are not as good at fitting the experimental data as well as predicting new data, compared to the titre models. This could be explained by the fact that HCP levels, in comparison to product titre, are not routinely monitored and optimised through mapping the design space of a production bioreactor.

**Table 5-5: Comparison between generated HCP models and HCP models found in literature. The comparison looks at the main parameters having a significant impact on product titre as well as R<sup>2</sup> and R<sup>2</sup> predicted**

HCP			
Literature comparison	Significant factors	R <sup>2</sup>	R <sup>2</sup> predicted
AMBR 1 (Low generation)	<ul style="list-style-type: none"> <li>• Temperature</li> </ul>	0.72	0.50
AMBR 2 (Mid generation)	<ul style="list-style-type: none"> <li>• pH</li> <li>• Feed start time</li> <li>• Media osmolality</li> </ul>	0.98	0.95
AMBR 3 (High generation)	<ul style="list-style-type: none"> <li>• Temperature</li> </ul>	0.66	0.55
Rouiller et al. (2012)	<ul style="list-style-type: none"> <li>• pH</li> <li>• DO</li> <li>• Culture duration</li> </ul>	0.92	0.88



**Figure 5-6: Comparison between observed and predicted values of AMBR 1 HCP. Runs 1-11 correspond to CS1 within AMBR 1, maintained at a temperature of 36.5 °C and runs 11-20 correspond to CS2 within AMBR 1, maintained at a temperature of 36.5 °C followed by a temperature shift to 33 °C in day 4 of culture**

The observed vs predicted graph shows that the model chosen gives a good prediction of experimental values.

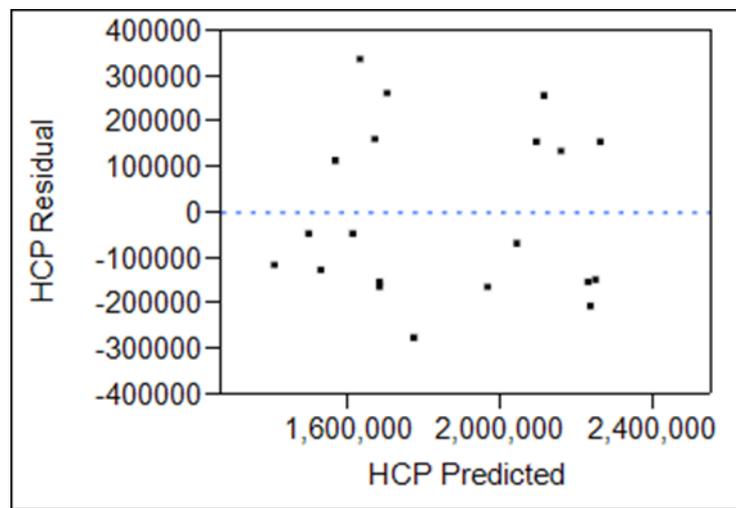
### 5.2.2.2 Verifying assumptions

In order to test the normality of residuals the Shapiro-Wilk test (as described in section 2.5.4.1) was performed. The Shapiro-Wilk statistics for HCP and its corresponding *p*-value are presented in Table 5-3. The *p*-value of 0.06 is not statistically significant therefore the hypothesis of normality of the residuals should not be rejected. This suggests that the assumption is not violated and the residuals are normally distributed.

To verify if the residuals are not correlated, the Durbin-Watson autocorrelation test (as described in section 2.5.4.2) was performed. The Durbin –

Watson statistics for HCP and its corresponding  $p$ -value are presented in Table 5-3. The  $p$ -value of 0.57 is not included within the two-tailed rejection region of  $p$ -value  $\leq 0.025$  or  $p$ -value  $\geq 0.975$ , therefore the null hypothesis of no autocorrelation is not rejected. The residuals are not correlated.

To check the linearity of the model, the residual vs predicted plot is evaluated (Fig. 5-7). The residuals are evenly distributed on either side of the regression line and there is no visible trend. The assumption of linearity is satisfied.



**Figure 5-7: Residual vs predicted values AMBR 1 HCP**

**Table 5-6: Final predictive models for titre and HCP from each AMBR run presented in coded factors. The coded factors A, B, C, D and E refer to the cell culture variables media osmolality, seeding density, temperature, pH and feed start time, respectively**

	Model	STATISTICS					
		R <sup>2</sup>	Predicted R <sup>2</sup>	Adjusted R <sup>2</sup>	R <sup>2</sup> k fold	RMSE	AIC
AMBR 1 – Titre	A*+B*+C*+AB*+BC	0.84	0.71	0.79	0.72	0.357	31.78
AMBR 1 – HCP	A+B+C*+A <sup>2</sup>	0.72	0.50	0.65	0.51	205141	558.72
AMBR 2 – Titre	B+D*+E*+BD*+DE*+D <sup>2</sup> *	0.98	0.97	0.98	0.97	0.289	25.65
AMBR 2 – HCP	B+D*+E+BE+B <sup>2</sup> *+D <sup>2</sup> *+E <sup>2</sup> *	0.98	0.95	0.97	0.95	125536	652.77
AMBR 3 – Titre	A+B*+C*+A <sup>2</sup> *	0.91	0.82	0.89	0.84	0.351	28.68
AMBR 3 – HCP	A+C*	0.66	0.55	0.62	0.57	206281	637.50

### **5.2.3 Importance of predictive models within QbD**

Following a meeting of the Process Analytical Technology Community of Practice of United Kingdom/Ireland, a survey aimed at finding out industrial experiences and opinions on the business benefits of Quality by Design was performed. The concept of modelling in QbD has numerous benefits and is widely used. All of the eleven companies surveyed (e.g. Abbott, Astra Zeneca, GSK, Merck, Pfizer, Eli Lilly, Bristol Myers Squibb, etc) are using Design of Experiments and empirical modelling. Modelling has been used for establishing an operating design space, predicting scale-up parameters, guiding process development and improving product and process understanding. The implementation of statistical and mathematical modelling in QbD as well as other elements within the QbD framework (process analytical technology (PAT) tools, critical quality attributes risk assessments, real time release testing, etc) have brought significant benefits to the pharmaceutical industry. Improved process, product knowledge and understanding, improved control strategy, improvement in product quality and product robustness/reproducibility, yield increase and cost reduction are just some of the main benefits (Kourti and Davis, 2012).

### **5.3 Conclusions**

This chapter describes the application of QbD principles (HT scale-down model combined with statistical modelling) to the cell culture process of an IgG1 monoclonal antibody, which resulted in the development of six predictive models (for low, mid and high generation cells). The experiments were performed using the ambr system and a DoE approach to identify the impact of cell culture inputs, both individually and as interactions, on critical quality attributes of the process (mAb titre and HCP levels). All of the parameters analysed (media osmolality, seeding density, culture temperature, culture pH and feed start time) were identified as main factors affecting mAb titre throughout all ambr runs, while culture temperature seemed to have the most significant impact on HCP levels, followed by culture pH, feed start time and media osmolality. The regression models generated were used to

explain the criticality of process parameters and allowed characterizing the impact of cell culture inputs on the performance of the cell culture process (antibody titre and HCP levels). The predictive models developed on the relationship between culture conditions, product titre and HCP levels will be used in the next chapter, within a whole process cost model, able to predict equipment sizes, cost of goods (COG) and optimal DSP process sequence associated with different cell culture strategies. This will enable rapid identification of the most promising and robust combinations of USP and DSP activities for more streamlined development in both existing and new facilities.

## Chapter 6

### 6 Integration of predictive cell culture correlations with bioprocess economics and uncertainty analysis

#### 6.1 Introduction

Chapter 3 and 5 focused on using high throughput experimentation and multivariate data analysis to derive predictive cause-and-effect correlations for cell culture. The correlations provide a link between operating parameters and the cell culture performance in terms of cell count, antibody titre and HCP levels. The next challenge was to link the correlations to the impact on downstream processing, the final drug substance characteristics and the cost of goods. This would enable overall process yields, purities and costs to be determined as function of the cell culture operating parameters. Commercially available process economics packages tend to use simple short-cut mass models for cell culture that assume the yield and purity *a priori* with no capacity to substitute user-defined models. Hence this chapter explores the potential of integrating the correlations with cost models so as to determine the overall performance and robustness of whole bioprocess strategies.

More specifically the aim of this chapter was to incorporate the predictive modelling equations generated in Chapter 5 with a prototype bioprocess economics and optimisation tool (Fig. 6-1, 6-2) in order to identify the most cost-effective combination of input cell culture conditions and chromatography column sizing strategies. The optimisation tool comprised a biomanufacturing process economics evaluation engine and database linked to a meta-heuristic optimisation algorithm. The impact of uncertainty in cell culture parameters on process performance and the likelihood of process metrics falling out of specification (output (kg), HCP<sub>final</sub> (ng/mg)) were assessed using Monte Carlo simulations.

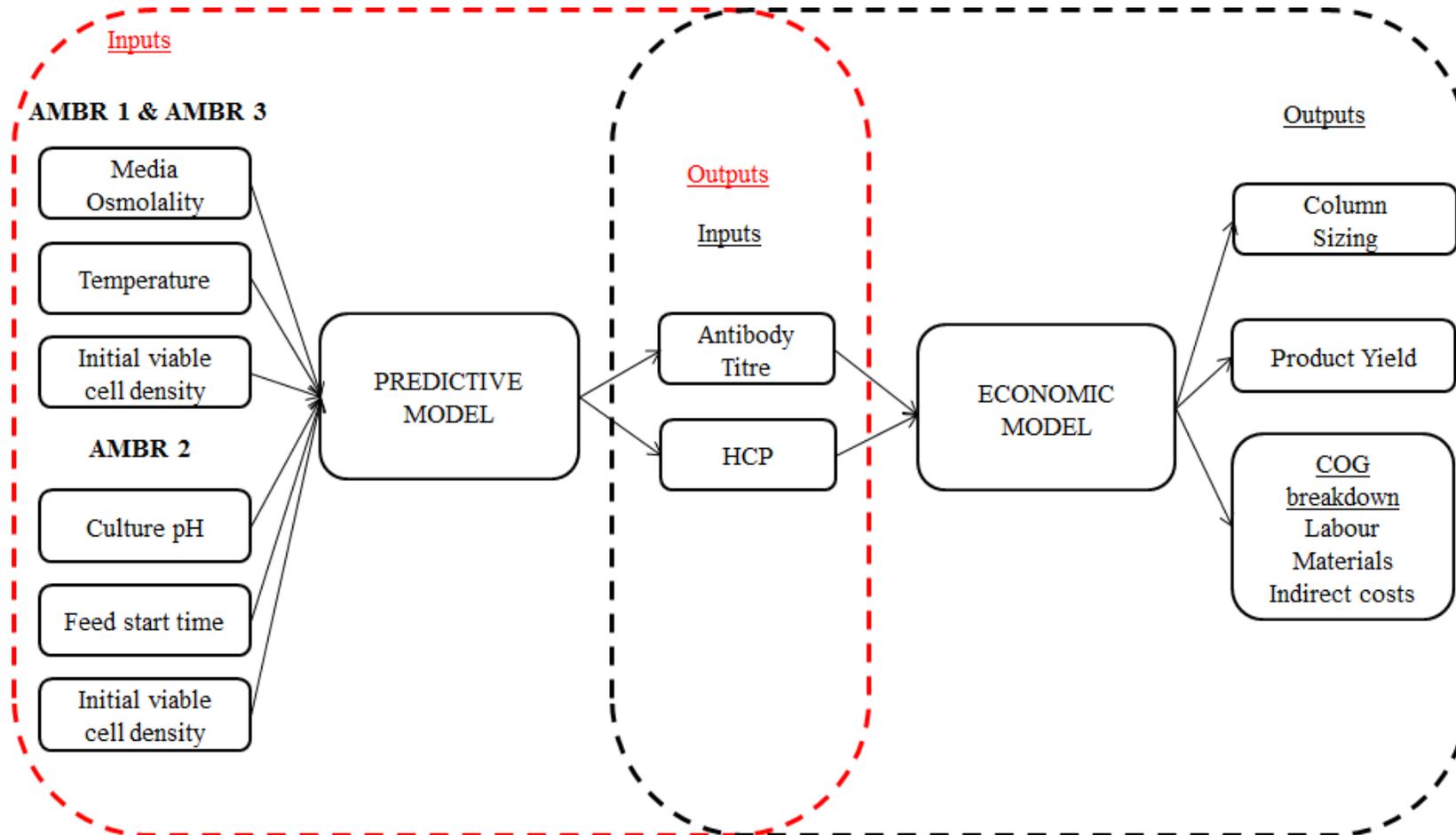


Figure 6-1: Overall integration of HT experimentation, predictive (statistical) modelling and economic modelling

## **6.2 Methodology**

### **6.2.1 Problem definition**

A bioprocess economics and optimisation tool using genetic algorithm developed at UCL (Simaria et al., 2012; Allmendinger et al., 2014) was firstly used to identify the impact of different titre and HCP loads from cell culture (resulting from AMBR 1, 2 and 3) on the COG of a fixed USP and DSP process and to identify the optimal sizing of a DSP process (measured by the lowest COG/g). Secondly, the tool was used to investigate the impact of uncertainty within cell culture input parameters on output (kg) and HCP<sub>final</sub> (ng/mg). In this chapter HCP is expressed as ng/mg, as opposed to ng/mL (Chapter 3). FDA's specification for HCP at the end of the manufacturing process is <100 ng/mg and as in this chapter, the impact of cell culture inputs on HCP<sub>final</sub> is assessed, this metric was chosen for HCP here. The optimisation problem considered here was subjected to an objective function, constraints and uncertain parameters, which are described in the following sections.

#### **6.2.1.1 Objective function**

The objective was to find the best combination of input parameters within each AMBR run that gives the minimal cost of goods per gram (COG/g) of manufactured product as well as understanding the impact of fluctuations in input parameters on the likelihood of process metrics falling out of specification.

#### **6.2.1.2 Constraints**

Various constraints need to be specified such that feasible results are generated by the optimisation tool.

#### *Demand and batch constraints*

This constraint ensures that the amount of product manufactured satisfies the annual demand specified. The estimated total number of batches should not surpass the maximum number of batches that fit within the facility, during a year.

#### *Purity constraints*

This constraint ensures that the final level of host cell proteins (HCP) meets the purity target. HCPs are impurities in a mAb manufacturing process that need to be closely monitored and removed during the purification process in order to ensure patient safety. Given the initial level of HCPs (measured in ng of HCP per mg of product), the DSP process is required to lower the final HCP level to the final product specification limit of < 100 ng/mg.

#### **6.2.1.3 Manufacturing uncertainties**

The manufacturing process of antibodies can be subject to several uncertain factors. Here the focus is on potential fluctuations within input cell culture parameters. Uncertainties in starting media osmolality and seeding density were considered for AMBR 1 and 3 while uncertainties in pH and seeding density were considered for AMBR 2. In order to account for fluctuations in input cell culture parameters, a triangular distribution around each parameter was used. The framework represents uncertainty by associating each factor with a probability distribution, from which values are drawn at random during Monte Carlo (MC) trials. Uncertainty related to cell culture parameters, introduces uncertainty within process metrics such as product titre and HCP levels.

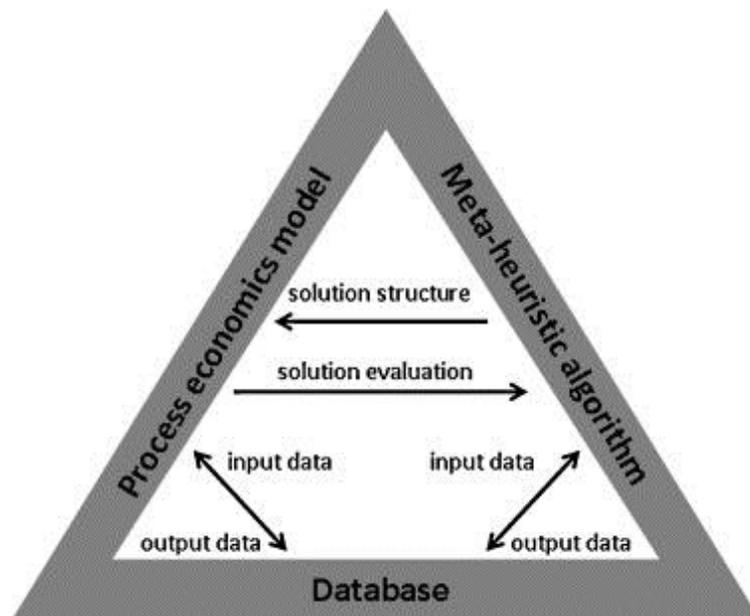
### **6.3 Framework description**

The tool consists of three main components, as presented in Fig. 6-2:

1. Meta-heuristic optimisation algorithm

2. Biomanufacturing process economic model
3. Database (Simaria et al., 2012)

The meta-heuristic optimisation algorithm searches the large decision space of process configurations and uses the process economic model to assess each alternative. The process economic model calculates several technical and financial outputs for a specific process configuration. The database keeps all the input data that needs to be used by the optimisation algorithm and process economics model, the output data that is generated by the framework as well as cost information with regards to resources such as labour and consumables.



**Figure 6-2: Main components of integrated bioprocess economics and optimisation tool\***

\* (Adapted from Simaria et al. (2012))

### **6.3.1 Bioprocess economics model**

The process economics model within the framework was based on previous UCL work (Lim et al., 2005; 2006 and Farid et al., 2005a; 2007b). The model was designed to use mass balance equations, time and detailed cost calculations in order

to determine cost of goods per gram (COG/g) of a specific process configuration. The COG contains both direct and indirect cost. The direct costs refer to the usage of resources such as materials (chromatography resins, buffers, membrane filters) and labour. The indirect costs are associated with the facility's maintenance costs, depreciation and general utilities. The COG/g value is then calculated by dividing COG by the total amount of product manufactured in a year (Simaria et al., 2012; Allmendinger et al., 2014a, 2014b).

## **6.4 Case study description**

An industrially-relevant case study was used to demonstrate the ability of the framework to calculate COG/g for different combinations of cell culture inputs and to determine the impact of uncertainty in the inputs on process outputs of a mAb manufacturing process (output (kg), HCP<sub>final</sub> (ng/mg)). The case study concentrates on a single-product mAb manufacturing facility that features a three chromatography step purification sequence to satisfy a total product demand of 500 kg/year. The chromatography step sequence was fixed to Protein A (Prot A) → Anion Exchange (AEX) → Cation Exchange (CEX). The first and the third steps use packed-bed chromatography while the second step is a membrane chromatography step. The fluctuations in input parameters were modelled using triangular distribution, as shown in Table 6-2.

## **6.5 Results and discussions**

### **6.5.1 Impact of cell culture parameters on COG/g**

The COG/g objective is mainly affected by variations in mAb titre and process yield (Allmendinger, 2014b). Process yield is calculated by adding together the individual step yields of each unit operation in the DSP process and then dividing by the number of DSP steps. The impact of cell culture parameters on COG/g of a fixed USP and DSP manufacturing process is evaluated here; therefore process yield is

unchanged and does not have an effect on COG/g in this case. The main factor influencing the COG/g is mAb titre, as it has a direct impact on the mass of product manufactured which is the denominator of the COG/g. Variations in mAb titre can occur due to variations in input cell culture parameters.

The model was updated to evaluate the impact of cell culture inputs used within the design of experiments for AMBR 1, 2 and 3 on mAb titre and HCP levels and consequently on COG/g. The experimental design is shown in Table 6-1. For each combination of cell culture inputs, the process economic model calculated the mAb titre and HCP levels using the predictive correlations generated in Chapter 5 and then used these cell culture outputs to determine process and economic metrics as well as column sizing strategies (Fig. 6-1).

**Table 6-1: The experimental design investigated for the input cell culture parameters within AMBR 1, 2 and 3**

	Design space			Number of combinations
	Minimum	Maximum	Increments	
<b>AMBR 1</b>				
Seeding density ( $\times 10^6$ cells/mL)	0.5	1	0.25	60
Media osmolality (mOsm/kg)	310	400	10	
Temperature ( $^{\circ}$ C)	33	36.5	N/A	
<b>AMBR 2</b>				
Seeding density ( $\times 10^6$ cells/mL)	0.5	1	0.25	27
pH	6.6	7	0.2	
Feed timing (h)	24	72	24	
Temperature ( $^{\circ}$ C)	36.5			
<b>AMBR 3</b>				
Seeding density ( $\times 10^6$ cells/mL)	0.6	1.8	0.4	72
Media osmolality (mOsm/kg)	310	390	10	
Temperature ( $^{\circ}$ C)	33	36.5	N/A	

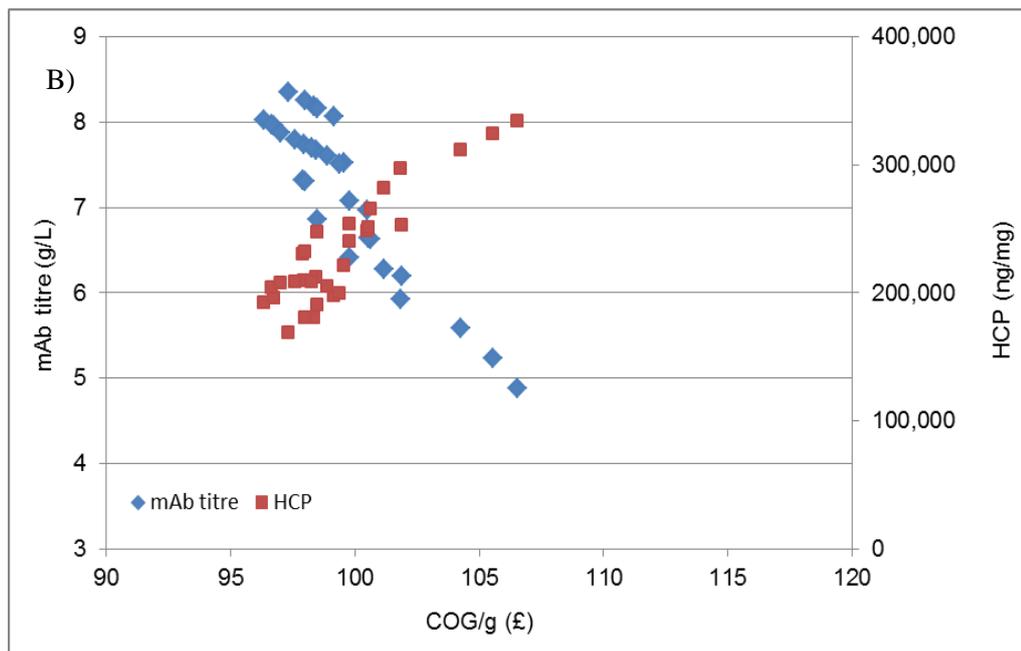
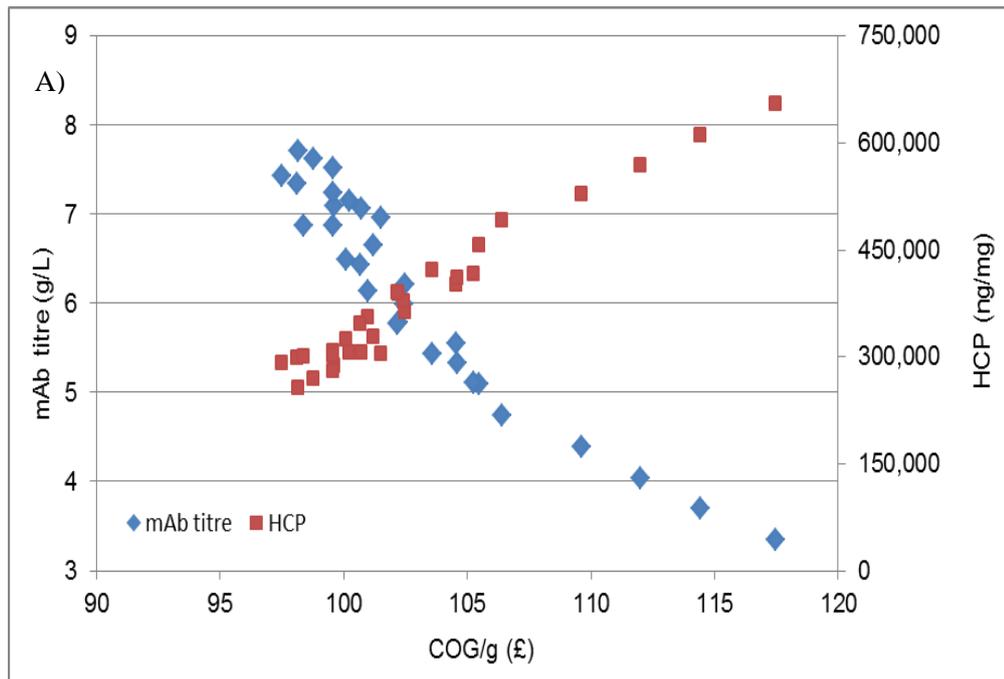
**Table 6-2: Details of case study scenario (Panel A) and probability distributions of uncertain factors (Panel B)\***

Panel A: case study scenario		
Parameter	Setting	
Chromatography steps	3	
Annual Demand	500 kg	
USP: DSP train ratio	2: 1	
Maximum final HCP level	100 ng/mg	
Bioreactor size	10,000 L	
Temperature	36.5 °C	
	AMBR 1/AMBR 3	AMBR 2
Timing of feed initiation	48 h	24 h
Panel B: probability distributions of uncertain factors		
Uncertain factor	Probability distribution*	Variation (%)
<b>AMBR 1</b>		
Seeding density ( $\times 10^6$ )	Tr (0.5, 0.75, 1)	33.3
Media osmolality	Tr (310, 350, 390)	11.4
<b>AMBR 2</b>		
Seeding density ( $\times 10^6$ )	Tr (0.5, 0.75, 1)	33.3
pH	Tr (6.6, 6.8, 7)	2.9
<b>AMBR 3</b>		
Seeding density ( $\times 10^6$ )	Tr (0.6, 1.2, 1.8)	50
Media osmolality	Tr (310, 350, 390)	11.4

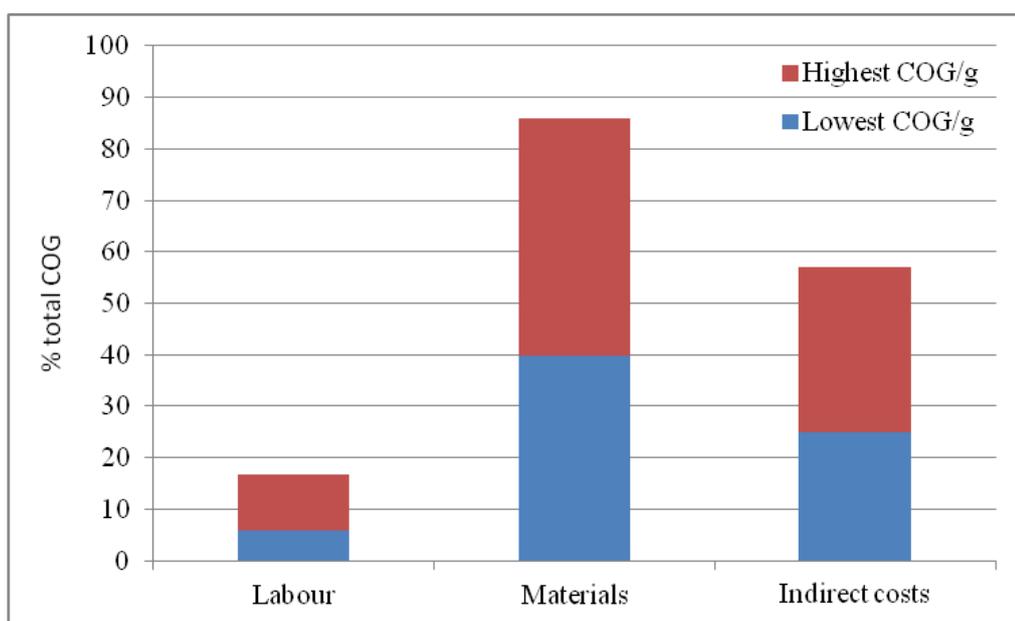
\* In order to account for fluctuations in input cell culture parameters, a triangular distribution around each parameter was used (Tr).

For AMBR 1, the impact of titre and HCP load resulting from the 30 different combinations of cell culture inputs, for each temperature, on COG/g is shown in Fig. 6-3. It can be seen that, as expected, COG/g is directly influenced by the mAb titre, with an increase in mAb titre resulting in a decrease in COG/g. For this set of experiments, in which low generation cells were used, it was previously shown that it is possible to increase antibody titre with little impact on HCP levels and hence subsequent DSP operations (Chapter 3). Here, it is shown that the combination of conditions that results in high mAb titre and low HCP levels also has the lowest COG/g. The combination of cell culture inputs, as well as the column sizing strategies for the lowest and highest COG/g achieved within each ambr run are presented in Table 6-3. The Protein A column sizing strategy for the lowest COG/g for each AMBR employs smaller columns (lower height) running for more cycles compared to the Protein A column sizing for the high COG/g cultures. This has a beneficial impact on the COG/g as it reduces the amount of Protein A resin used. Protein A is one of the most expensive unit operations within the manufacturing process of therapeutic proteins, particularly because of its highly expensive resin; therefore any savings in the amount of resin used would reduce the COG/g.

A comparison between the COG breakdown for the lowest and highest COG/g achieved in AMBR 1 is presented in Fig. 6-4. The direct costs include the usage of resources such as materials (chromatography resins, membrane filters, buffers) and labour while the indirect costs are associated with the facility's maintenance costs, depreciation and general utilities. A demand of 500 kg per year was assumed for this case study and in order for the lower titre cultures to be able to satisfy the annual demand; a higher number of batches is required. This will have an impact on the cost of labour, as a higher number of personnel would be necessary. The highest COG/g was achieved within cultures with the lowest mAb titre (Fig. 6-3) which explains the higher labour costs for these cultures. In higher titre cultures (lower COG/g) the resin utilisation is higher compared to lower titre cultures (72 % vs 63 %) which resulted in a lower cost of consumables, due to the fact that the cost of resins is a major contributor to the overall cost of consumables.



**Figure 6-3: The impact of titre and HCP load resulting from AMBR 1 cell culture at A) 36.5 °C and B) 33 °C on the COG/g of a fixed USP and DSP process.**



**Figure 6-4: COG breakdown for the lowest and highest COG/g achieved within the design space of AMBR 1**

For AMBR 2, the impact of cell culture inputs on mAb and HCP levels is shown in Fig. 6-5 and on COG/g in Fig. 6-6. It can be seen that pH 6.6 has a significant negative impact on mAb titre and HCP levels, compared to pH 6.8 and 7 (Fig. 6-5). This is due to its impact on suppressing cell growth (Tsao et al., 2005; Trummer et al., 2006). This negative impact on mAb titre, results in a significant increase in COG/g, regardless of other cell culture conditions within the DoE design (Fig. 6-6). Within a constant pH (6.8 and 7) a delay in the timing of feed initiation from 24 h to 72 h, resulted in a decrease in mAb titre and HCP levels for  $0.5$  and  $0.75 \times 10^6$  cells/mL seeding density, while for  $1 \times 10^6$  cells/mL seeding density, a delay in timing of feed initiation resulted in a decrease in mAb titre and an increase in HCP levels. Generally, a delay in timing of feed initiation, led to an increase in COG/g, except for cultures with a pH of 6.6.

**Table 6-3: Combination of input cell culture parameters for AMBR 1, 2 and 3 with their associated Prot A and CEX column sizing, for the lowest and highest COG/g achieved in each case**

	Combination of cell culture inputs					COG/g (£)	Column sizing	
	Media Osmolality (mOsm/kg)	Seeding Density ( $\times 10^6$ cells/mL)	Temperature ( $^{\circ}$ C)	pH	Time of feed initiation (h)		Prot A	CEX
							height – diameter – cycles (cm – cm – no)	
AMBR 1	310	1	36.5	6.8	48	96.3	20-120-7	24-90-9
	400	0.5	33	6.8	48	117.5	24-100-3	19-120-3
AMBR 2	310	0.5	36.5	7	24	99.4	23-100-9	24-120-5
	310	0.5	36.5	6.6	24	194.3	23-120-1	17-90-2
AMBR 3	370	1.8	33	6.8	48	96.4	23-120-8	15-120-10
	390	0.6	36.5	6.8	48	103.4	24-100-7	19-90-9

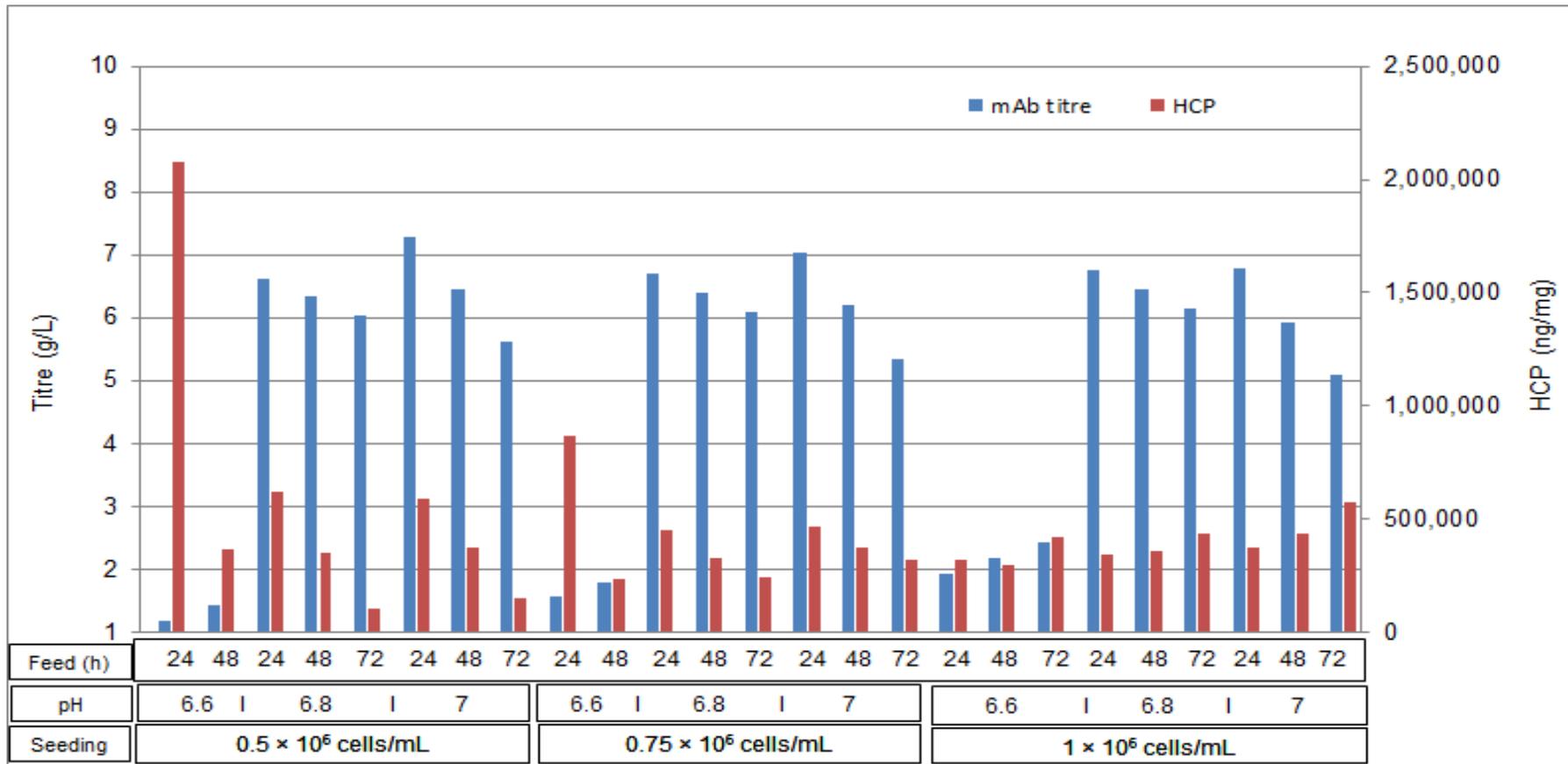


Figure 6-5: The impact of cell culture inputs on mAb titre and HCP levels within AMBR 2 cultures

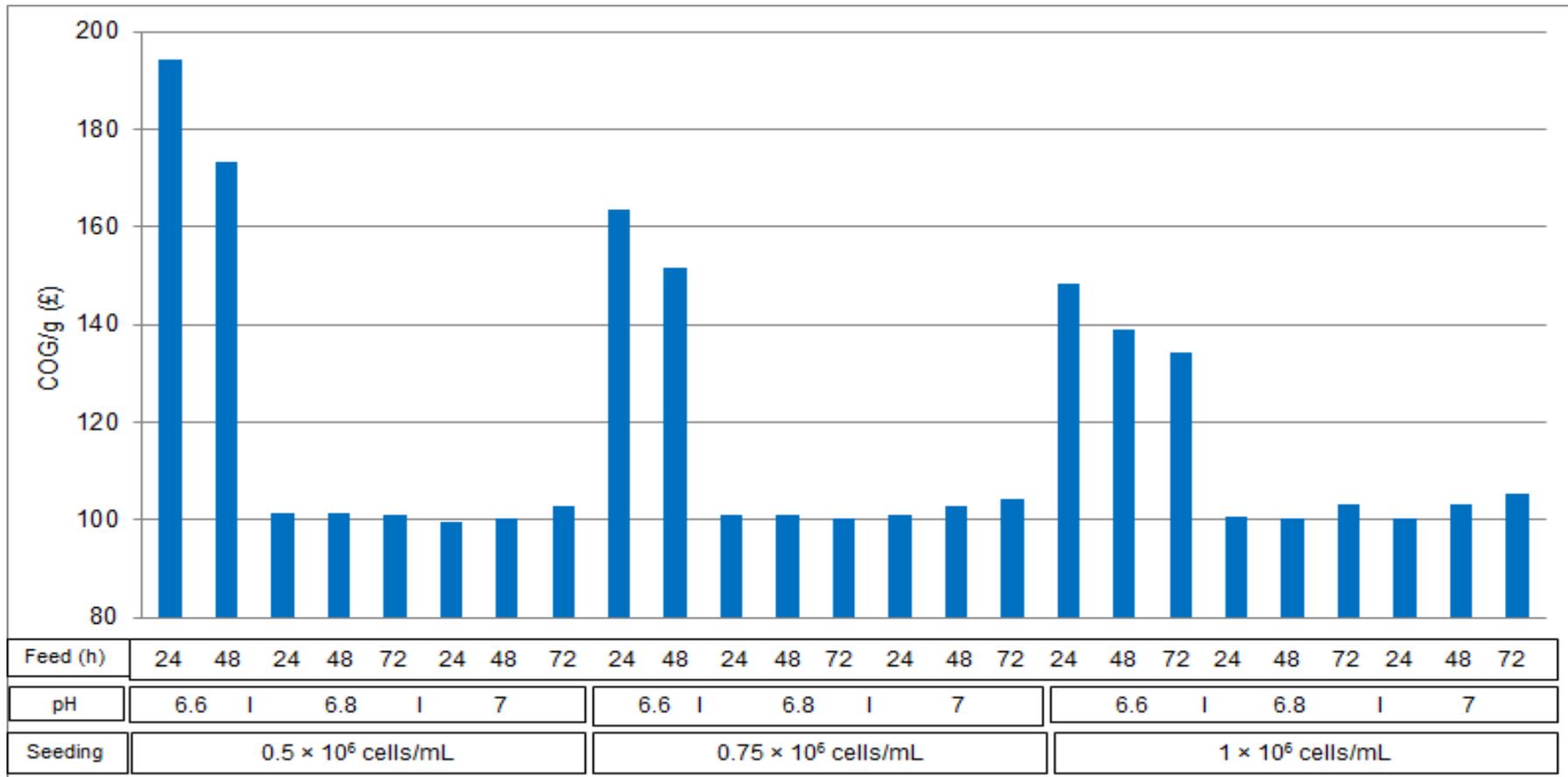


Figure 6-6: The impact of cell culture inputs on COG/g within AMBR 2 cultures

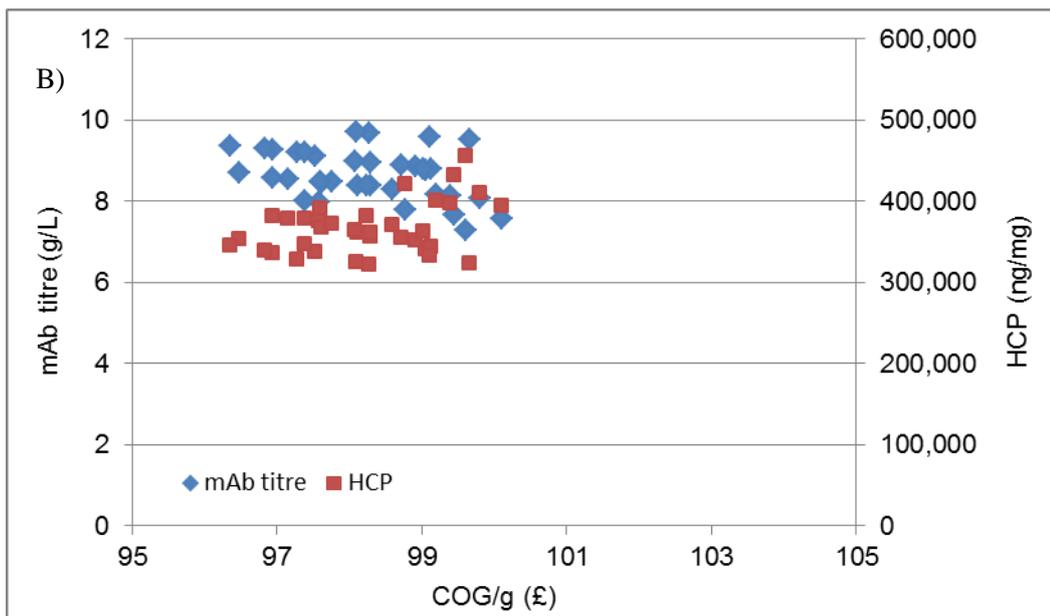
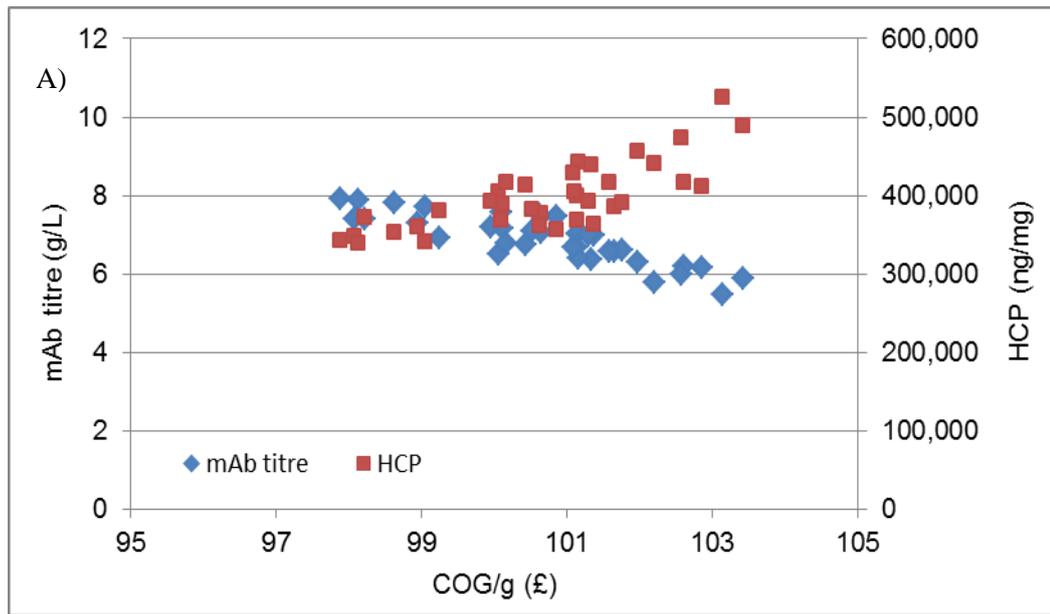
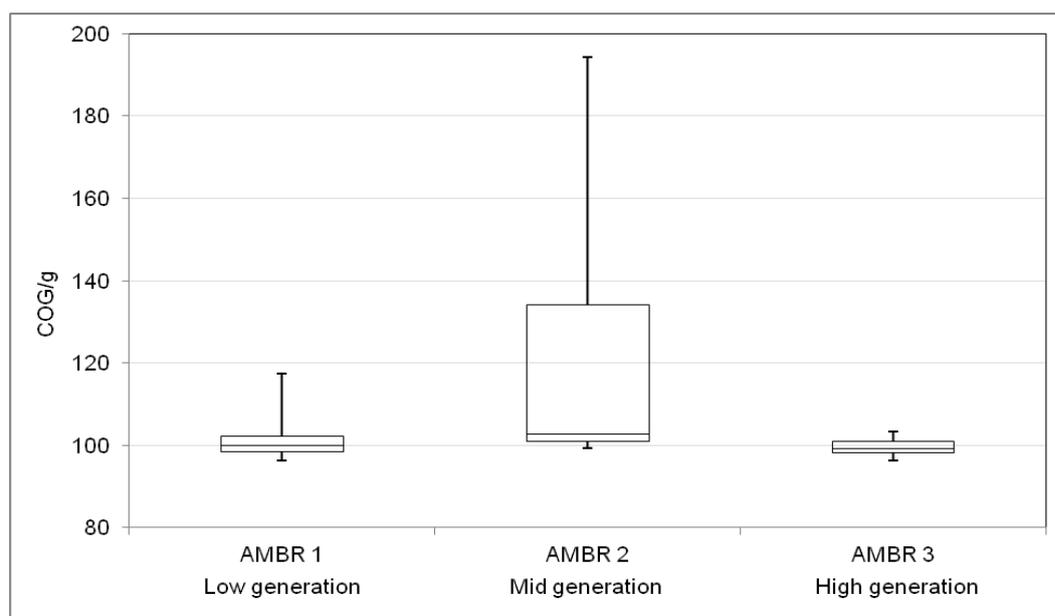


Figure 6-7: The impact of titre and HCP load resulting from AMBR 3 cell culture at A) 36.5 °C and B) 33 °C on the COG/g of a fixed USP and DSP process.

For AMBR 3, the impact of titre and HCP load resulting from the 36 different combinations of cell culture inputs on COG/g is shown in Fig. 6-7. Here, the COG/g is again directly influenced by the mAb titre, an increase in mAb titre resulting in a decrease in COG/g. This observation is more apparent in 36.5 °C cultures.



**Figure 6-8: Boxplots showing the distribution of COG/g for different AMBR runs. The box represents the 25<sup>th</sup> and 75<sup>th</sup> percentile with the median indicated by the middle horizontal line. The whiskers represent the observations with the lowest and highest COG/g for each AMBR run.**

The distribution of COG/g for all the combinations in cell culture inputs in each AMBR run is presented using box plots in Fig. 6-8. AMBR 2 (using mid generation cells) displayed the widest variation in COG/g. This is due to pH 6.6 which resulted in significantly lower titres than pH 6.8 and 7, resulting in high COG/g.

### 6.5.1.1 Monte Carlo simulations

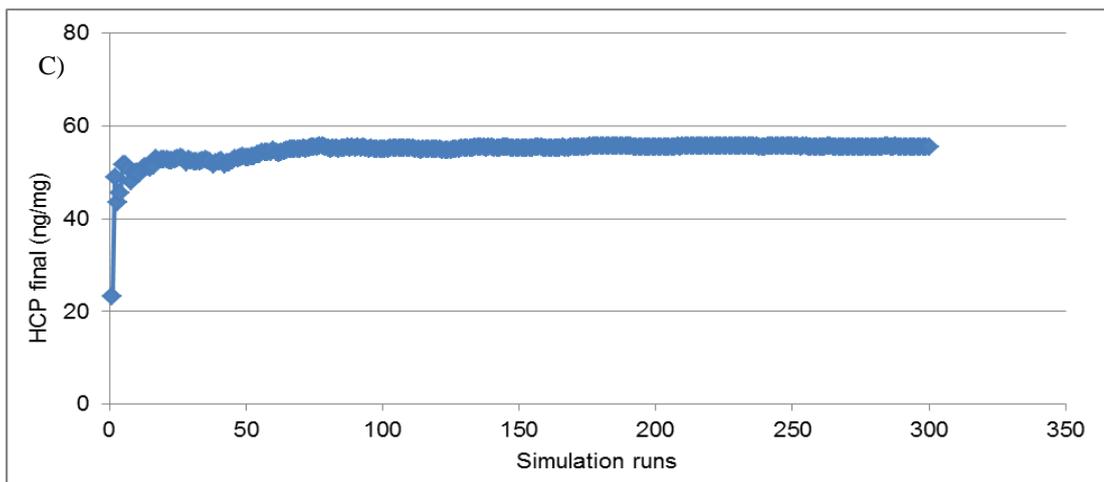
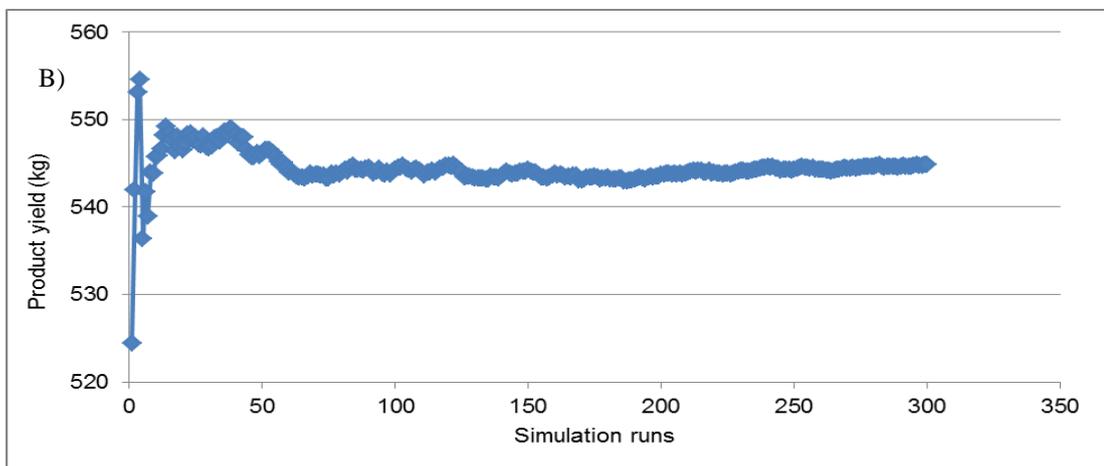
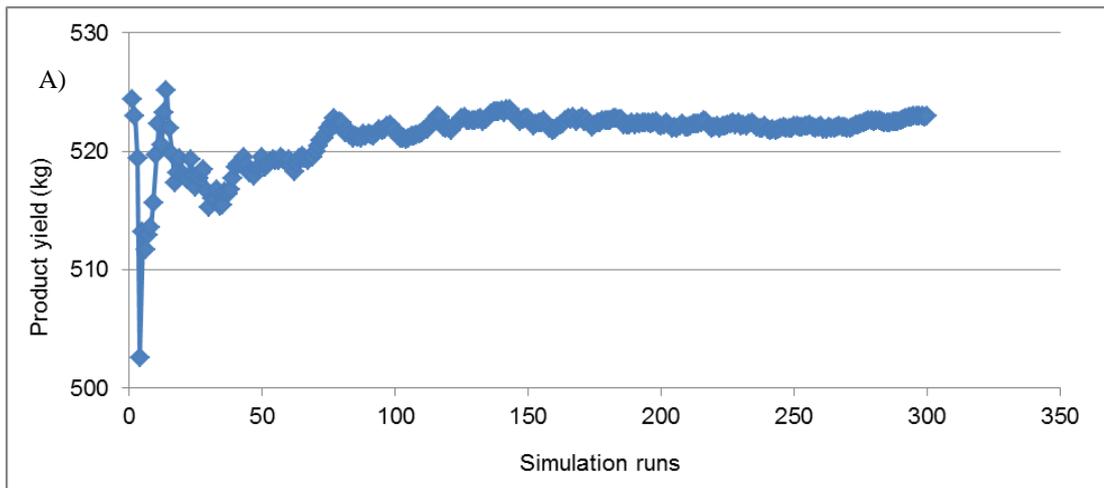
In order to determine the number of simulation runs required to reach convergence, running averages of the simulation results were monitored until they levelled off (Fig. 6-9). This translated into running 150 Monte Carlo simulation runs in order to reach convergence in the output metrics and to characterise the variability in

performance measures (output and HCP<sub>final</sub>) due to uncertainty in seeding density and media osmolality for AMBR 1 and 3, and seeding density and pH for AMBR 2.

### **6.5.1.2 Analysis**

The tool was used to predict the likelihood of product loss or failure to meet demand in AMBR 1 and 3, while in AMBR 2 the likelihood of the final HCP level (HCP<sub>final</sub>) being higher than 70 ng/ mg was assessed. MC simulations were employed to generate possible random outcomes by using the probability distribution of the input parameters, leading to a frequency distribution of outcomes for each output (output (kg), HCP<sub>final</sub> (ng/mg)). From these distributions the likelihood of particular thresholds being achieved can be determined (Stonier et al., 2013).

The expected fluctuations in cell culture input parameters can result in fluctuations in product titre and final HCP. Uncertainty within product titre can have a considerable impact on output. Fluctuations in product titre can lead to a) failure to meet demand (if fluctuation in cell culture parameters result in a lower titre than expected) or b) product waste (if resulting titre is higher than expected and the DSP cannot handle the excess) (Allmendinger, 2014a). Uncertainty with regards to processing time and step yields can also exist within a manufacturing process but are not examined in this thesis.



**Figure 6-9: Running average in product yield for A) AMBR 1, C) AMBR 3 and in HCP final for B) AMBR 2 during Monte Carlo simulation trials.**

For AMBR 1 and 3 a product output in the range of 500-550 kg/year was considered acceptable, with anything below or over this range, being considered out of specification. For AMBR 2,  $HCP_{final} < 70$  ng/mg was considered in specification and anything above it out of specification. The general guidelines for HCP levels in the final product are set to less than 100 ppm (ng/mg). There are currently no guidelines referring to levels of specific HCPs within the whole population that can be found in the final product and have the potential to impact on the product structural stability. Bracewell et al. (2015) questions if the final HCP limit of 100 ng/mg is acceptable and emphasizes that there should be more detailed criteria, based on a better understanding of the HCP population present. In all three sets of experiments performed, all of the cell culture input combinations satisfied the 100 ng/mg limit of HCPs in the final product. Considering that the 100 ng/mg limit might not be acceptable in the future, within AMBR 2 the probability of  $HCP_{final} > 70$  ng/mg was considered as being out of specification.

The probability distributions (based on triangular distribution) used for cell culture input factors within each AMBR, with their corresponding % variation are shown in Table 6-2. These distributions are the initial distributions tested. Density plots showing the probability of falling out of specification with respect to output are shown in Fig. 6-11 A), B), C) for AMBR 1 and in Fig. 6-11 D), E), F) for AMBR 3. Different changes made to the initial distributions were also assessed. This allowed the identification of how tight cell culture inputs need to be controlled for the product output and  $HCP_{final}$  not to fall out of specification.

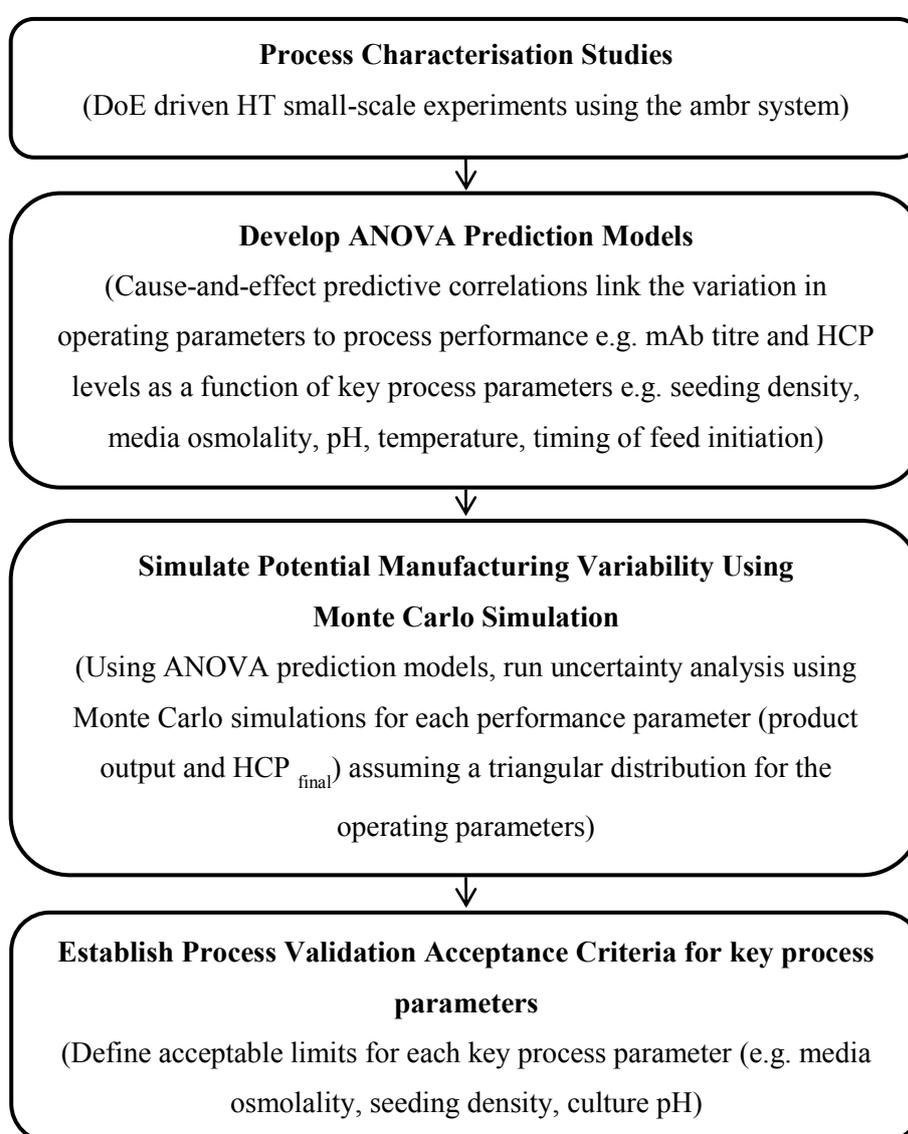
In AMBR 1 experiments, using the initial probability distributions, a 50 % probability of failing to meet demand and a 1 % probability of product waste were identified (Fig. 6-11 A)). In Chapter 3 it was shown that at 36.5 °C, mAb titre within AMBR 1, was influenced more by media osmolality than seeding density (Fig. 3-2). Considering this, the probability distribution for seeding density was kept constant while the distribution for media osmolality was first tightened to  $340 \pm 30$  (310 – 370 mOsm/kg). In this case, the probability of failing to meet demand was lowered to 25

% while the probability for product loss increased to 19 % (Fig. 6-10 B)). Product loss can occur in situations when titres are higher than expected and the chromatography columns do not have sufficient excess capacity to cope with higher product loads. Subsequently, the media osmolality distribution was tightened even further to  $330 \pm 20$  (310 – 350 mOsm/kg) and the probability for both failure to meet demand and product lost were within acceptable limits (Fig. 6-11 C)). A 10% probability of failure was considered as being acceptable. A tighter control in the starting media osmolality needs to be performed in order to maintain process robustness.

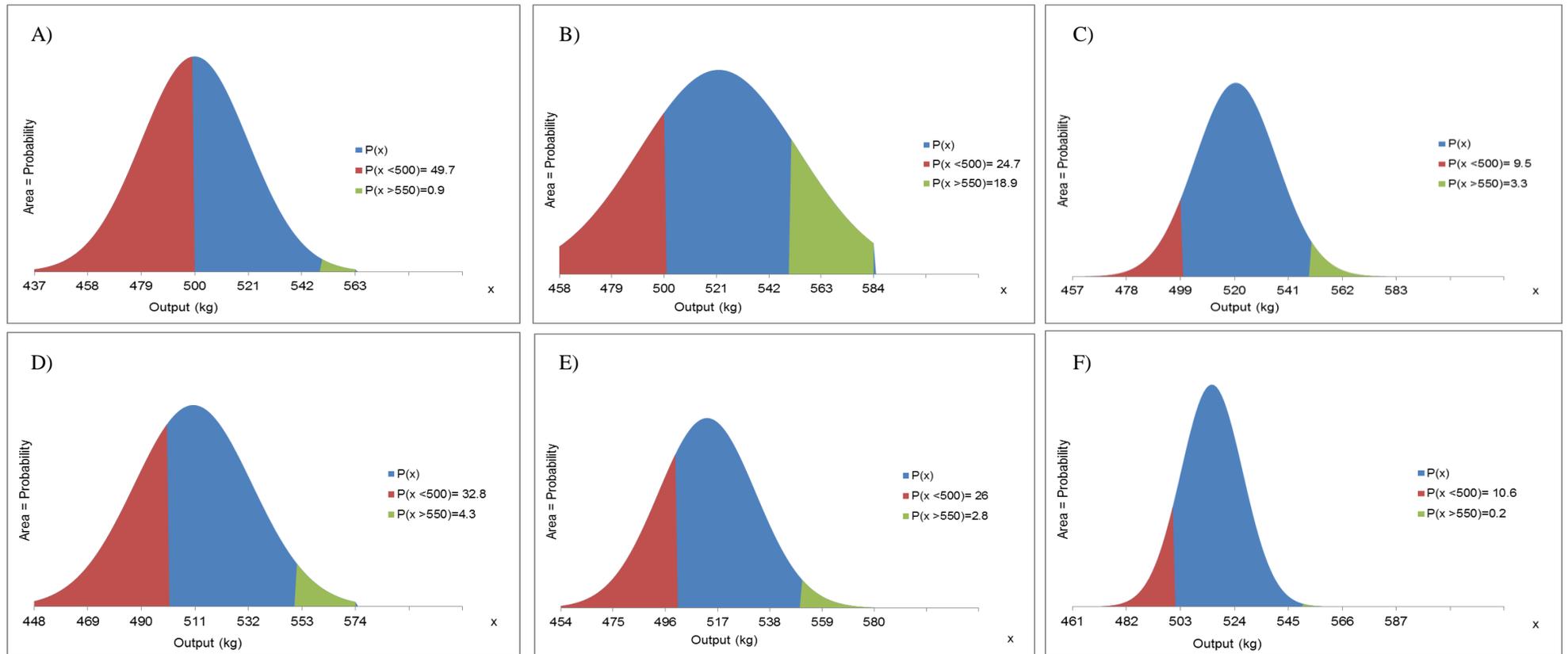
In AMBR 2 experiments, probability distribution within input parameters was used to assess the probability of HCP<sub>final</sub> (ng/mg) falling out of spec (HCP<sub>final</sub> > 70 ng/mg). Using the initial probability distributions (Table 6-2), there is a 67 % probability of HCP<sub>final</sub> being higher than 70 ng/mg (Fig. 6.12 A)). It has been shown that pH has a significant impact on growth, mAb titre and HCP levels therefore a tighter pH control ( $6.8 \pm 0.1$ ) was assessed. This resulted in a decrease in the probability to 14 % (Fig. 6-12 B)). Cultures at pH 6.8, within AMBR 2, showed lower levels of HCP at harvest compared to cultures at pH 7 (Fig. 3-2). It was anticipated that a tighter pH control ( $6.8 \pm 0.1$ ) would result in lower levels of HCP<sub>final</sub>, for a fixed USP and DSP process, compared to a pH control of  $6.8 \pm 0.2$ . A further change was made in the probability distribution for seeding density ( $0.75 \pm 0.2 \times 10^6$  cells/mL) resulting in 0% chance of HCP<sub>final</sub> being higher than 70 ng/mg (Fig. 6-12 C)). Culture pH needs to be controlled within a tight range to ensure that the final HCP levels do not exceed a certain limit.

For AMBR 3, using the initial probability distributions, a 33 % probability of failing to meet demand and a 4 % probability of product waste were identified (Fig. 6-11 D)). Lowering the osmolality distribution range from  $350 \pm 40$  (310 – 390 mOsm/kg) to  $340 \pm 30$  (310 – 370 mOsm/kg) resulted in an improvement in the probability for both ends of the spectrum (Fig. 6-11 E)). After exploring several changes within the seeding and osmolality distribution ranges, it was found that for a

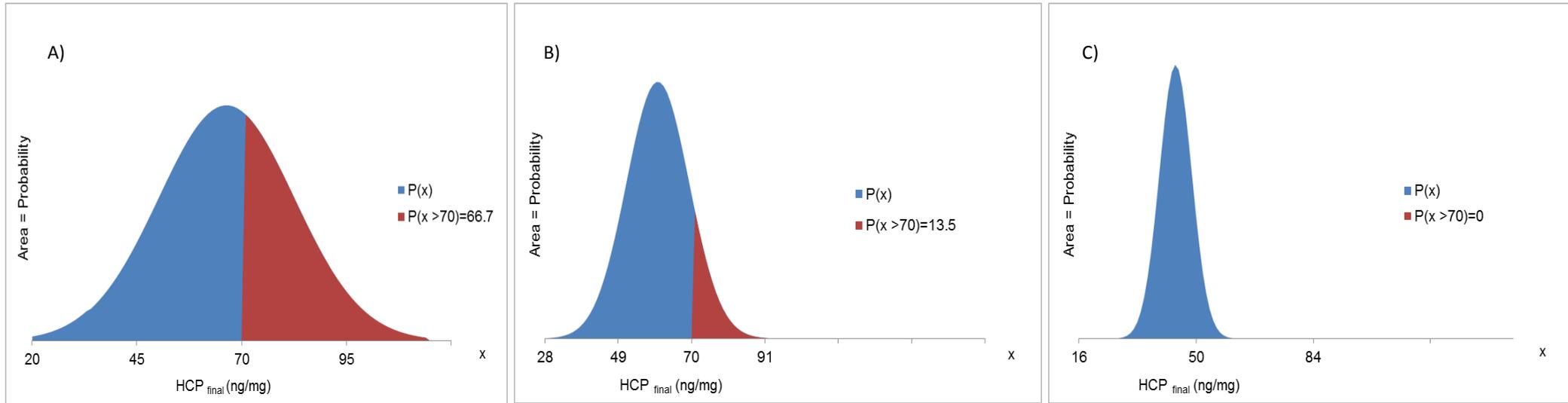
seeding density of  $1.2 \pm 0.2 \times 10^6$  cells/mL and an osmolality of  $340 \pm 30$  mOsm/kg, the probability of falling out of specification is relatively close to the acceptable limits (Fig. 6-11 F)). As for AMBR 1, a tighter control in starting media osmolality is required to ensure process robustness. Performing uncertainty studies using MC simulations, combined with predictive correlations can be used to determine process validation acceptance criteria (PVAC) for key process parameters (media osmolality, seeding density, pH) (Fig. 6-10). Similar studies have been performed by Looby et al. (2011) and Gommers et al. (2014).



**Figure 6-10: Steps used for establishing PVAC using HT experimentation, ANOVA and Monte Carlo simulation methods**



**Figure 6-11: Density plots showing the impact of tighter osmolality and seeding density control on the probability ( $P(x)$ ) of the product output falling out of specification in AMBR 1 and AMBR 3, using MC = 150 for Seeding density  $0.75 \pm 0.25 \times 10^6$  cells/mL A) Osmolality  $350 \pm 40$  mOsm/kg; B) Osmolality  $340 \pm 30$  mOsm/kg; C) Osmolality  $330 \pm 20$  mOsm/kg in AMBR 1 and for Seeding density  $1.2 \pm 0.6 \times 10^6$  cells/mL D) Osmolality  $350 \pm 40$  mOsm/kg; E) Osmolality  $340 \pm 30$  mOsm/kg; F) Seeding density  $1.2 \pm 0.2 \times 10^6$  cells/mL, Osmolality  $340 \pm 30$  mOsm/kg in AMBR 3**



**Figure 6-12: Density plots showing the impact of tighter pH and seeding density control on the probability (P(x)) of the HCP<sub>final</sub> (ng/mg) falling out of specification in AMBR 2, using MC = 150 for Seeding density  $0.75 \pm 0.25 \times 10^6$  cells/mL A) pH  $6.8 \pm 0.2$ ; B) pH  $6.8 \pm 0.1$ ; C) Seeding density  $0.75 \pm 0.2 \times 10^6$  cells/mL, pH  $6.8 \pm 0.1$**

## 6.6 Conclusions

The application of a meta-heuristic optimisation tool using genetic algorithms to an industrially relevant case study allowed the identification of the most cost-effective combinations of input cell culture conditions and column sizing strategies for AMBR 1, 2 and 3 (low, medium and high generation number cultures). The stochastic aspect of the tool was used to perform an uncertainty analysis, using Monte Carlo simulations, to help understand the impact of uncertainty in cell culture inputs on process performance. Antibody titre was identified as the main factor having an impact on COG/g, an increase in titre resulting in a decrease in COG/g. Cell culture inputs that resulted in higher titres, also resulted in lower COG/g. Under the current model assumptions, at high mAb titres, the sizing strategy for Protein A chromatography employs smaller columns with a higher number of cycles, saving on the quantities of resin used and in this case leading to a lower COG/g. This is dependent on the resin/membrane reuse strategy. Within the low generation cell culture (AMBR 1), a tighter control of the media osmolality range ensured that the resulting product output did not fail to meet demand or allowed for product waste, within the acceptable limits. In the higher generation cell culture (AMBR 3), a tighter control in both seeding density and media osmolality ranges was necessary to remain in specification. These uncertainty studies using MC simulations linked to ANOVA derived cause-and-effect predictive correlations can be used to establish acceptable operating ranges for key process parameters (media osmolality, seeding density, pH).

The results can be used by the cell culture process development team within a biopharmaceutical company to understand which cell culture parameters need a tighter control in order to avoid process outputs falling out of specification. Using the integrated QbD framework (process characterization studies based on DoE experiments using the ambr system, statistical analysis for deriving predictive correlations and uncertainty analysis using MC simulations), proven acceptable ranges (PARs)/process validation acceptance criteria (PVAC) for key process parameters can be determined (Fig. 6-10). Systematic approaches similar to the one

showed in this thesis are likely to be the core of “QbD-based” future regulatory submissions (Looby et al., 2011).

## Chapter 7

### 7 Conclusions and Future work

#### 7.1 Conclusions

The overall output of this research is a systematic framework combining state-of-the-art high throughput cell culture experiments (ambr system) with statistical correlations (derived through multivariate analysis techniques) and process economic models. This integrated framework enabled the identification of cell culture strategies that balance the needs of upstream and downstream manufacturability, robustness to process variations and cost-effectiveness, early in the development cycle.

DoE designed experiments were performed using the ambr system in order to evaluate the impact, different increases in antibody titres has on the resulting host cell protein levels at harvest. It has been shown that increases in mAb titre can be a result of different combinations of increases in cell densities and specific cell productivities (Fig. 3-3), each with a different impact on HCP levels and subsequent DSP steps. A Quality by design approach to cell culture process development was used to demonstrate that there is scope for cell culture processes in which the ratio of mAb to HCP can be increased and the association of mAb titre to HCP reduced. A combination of cell culture parameters that resulted in an increase of antibody titre, without a subsequent increase in HCP levels was identified within AMBR 1 experiments (36.5 °C, 313 mOsm kg<sup>-1</sup> media osmolality, 1 × 10<sup>6</sup> cells mL<sup>-1</sup> seeding density, pH 6.8 and low cell generation number). Within these experiments, three different cell generation cells were used and a stability issue was observed within higher cell generation cultures at 36.5 °C while at a lower temperature of 33 °C cell line instability was not apparent. Cells of a higher generation lead to an elevated ratio of HCP to product (AMBR 2, 3) as compared to lower generation cells (AMBR 1).

Recent literature has emphasized the need for a better understanding of the impurity profile entering DSP and the cell culture factors that are likely to change this profile, as certain problematic HCPs have been identified in the final drug product (Aboulaich et al., 2014; Thompson et al., 2014). There is a limited understanding regarding the relationship between the protein of interest, bioprocess conditions (both in USP and DSP) and problematic HCPs (Bracewell and Smales, 2013). In addition to assessing the impact of cell culture inputs on the resulting mAb titre and HCP levels at harvest, a deeper understanding was provided on the impact of cell culture parameters on the activity of specific problematic HCPs (proteases). Proteases are a sub-class of host cell proteins, identified as problematic by a variety of publications, due to their ability of associating with the product throughout the DSP process or having an impact on the product's structural stability (Dorai and Ganguly, 2014; Aboulaich et al., 2014; Wang et al., 2014). Culture temperature was identified as having the most significant impact on protease activity at harvest, a decrease in temperature resulting in a decrease in protease activity. Apart from temperature, other cell culture conditions investigated (seeding density and media osmolality) did not seem to have a considerable impact on protease activity. The relationship between protease and HCP levels was also assessed and it was shown that an increase in HCP levels does not result in a similar increase in protease activity.

The high throughput cell culture data generate using the ambr system was characterised by multivariate data analysis techniques (multiple linear regression, all possible stepwise regression) to derive statistical cause-and-effect correlations. These statistical equations are able to predict cell culture outputs such as mAb titre and HCP levels at harvest, based on cell culture inputs (temperature, media osmolality, seeding density, pH and timing of feed initiation). A set of equations for product titre and HCP levels were derived from AMBR 1, 2 and 3 (each set of experiments presenting a different cell generation number).

The resulting cell culture predictive correlations were linked to a meta-heuristic optimisation framework (optimisation algorithm, a biomanufacturing process economic model and a database). This allowed the selection of the most cost-effective combination of input cell culture conditions and chromatography column sizing strategies (Table 6-3). The impact of uncertainty in cell culture parameters on process performance and the likelihood of process metrics falling out of specification were assessed using Monte Carlo simulations. The COG/g objective function was significantly influenced by mAb titre, an increase in titre resulting in lower COG/g. Within AMBR 1 and AMBR 3 the variability in product yield (kg) due to uncertainty in seeding density and media osmolality was examined, while for AMBR 2, the variability in final HCP (ng/mg) due to uncertainty in seeding density and pH was assessed. Media osmolality and culture pH were identified as critical culture parameters, that need to be closely monitored in order to ensure that the resulting process metrics do not fall out of specification. Fluctuations in input parameters that can cause fluctuations in product titre can lead to a situation in which the demand is not met or expensive product is wasted due to the insufficient capacity of purification steps to handle the increased product load.

## **7.2 Future work**

There are a few areas within this work that might benefit from additional investigation or alternative approaches. Model validation is necessary in order to assess the ability of a statistical model to accurately predict outputs. This can be tested through internal and external validation. As previously presented in Section 5.2.1.1, internal validation (using k-fold cross validation) has been used as one of the criteria to help identify the model with the best prediction capability. To further investigate how good the chosen models are at predicting outputs with new combinations of input variables (within the same design space), external validation should be performed. Future work should involve setting up an experimental design with new combinations of input variables. Once the experiments are completed, the measured output values (e.g. titre and HCP) should be compared with the previously

predicted values in order to evaluate how efficient the model is at externally predicting outputs.

The systematic framework combining microscale experimentation with statistical correlations and process economic models has been shown to be an effective way of identifying robust and cost-effective cell culture strategies. This framework provides a foundation to build upon, where predictive correlations can be created for downstream processing steps such as primary recovery and chromatography. High throughput scale-down experimentation can be adopted to generate data in order to characterise the impact of cell culture conditions on clarification and purification potential. This data can then be used to develop models that link cell culture properties (e.g. cell density, viability, mAb titres) and recovery operating conditions (e.g. flow rate) to the performance of centrifugation and depth filtration (e.g. clarification, yield, purity, filter capacity) and purification potential.

The impact of different ways of increasing antibody titre on HCP levels at harvest has been assessed within a high-producing CHO cell line. Three sets of experiments, each using a different generation of cells, resulted in three different bioreactor scenarios. The cell line used was seen to exhibit instability, with lower specificity for the mAb ( $q_{\text{Mab}}$ ) in higher generation cultures. It is uncertain if the three scenarios identified are caused by the significant influence of the generation number on the cultures or by the different combinations of conditions within the three DoE designs examined. The application of the QbD approach (ambr system and Gyrolab workstation high throughput platform, linked with DoE) to a high producing stable CHO cell line could help identify the underlying cause of the different scenarios seen. If a cell line is seen to display a scenario in which the increase in antibody titres would not result in a subsequent increase in HCP levels, the optimisation strategies should be focused around the cell culture parameters that resulted in such a trend.

Cell culture processes are very complex unit operations and they generally require a large number of factors to be characterised in order to be able to try and accurately describe such processes. In the experiments performed, cell culture factors

such as temperature, seeding density, media osmolality, pH and timing of feed initiation were used to characterise cell culture processes. Further work can be focused on assessing the impact of additional cell culture parameters such as levels of dissolved oxygen, agitation speed as well as additional culture temperatures over a wider range, on antibody titres and HCP levels at harvest. Different outputs such as aggregate levels and DNA could also be investigated.

The antibody sector has seen significant increases in upstream productivities in recent years, with titres reaching over  $10 \text{ g L}^{-1}$  in fed-batch cultures. This increase in USP titres has resulted in increased focus to be placed on how these USP changes impact on the HCP profile and levels following onto the downstream process. It has been reported that certain HCP species can be found in the final drug product (Aboulaich et al., 2014; Thompson et al., 2014). Recent studies have confirmed the need for a better understanding of the presence and potential risk of certain problematic HCPs. In this work the impact of cell culture inputs on protease activity (identified as problematic HCPs) was assessed as well as the relationship between protease activity and total HCP levels. It was shown that the protease activity does not increase with an increase in HCP levels. Other published work, looked at the relationship between other problematic HCPs and total HCP levels (e.g. PLBL2, Yuk et al., 2015) and showed that an increase in PLBL2 concentration was seen with an increase in HCP concentration. This suggests that different species within the whole HCP population are expressed differently in relation to the overall HCP trend. Future work could be focused on investigating the impact of cell culture conditions on the levels of other problematic HCPs and evaluate how their levels change with the overall HCP levels. Identifying cell culture inputs that have an effect on the levels of different problematic HCPs could help design processes that not only result in lower total HCP levels but also lower levels of specific, problematic HCPs and in this way lowering the burden on to DSP.

As previously mentioned, one of the limitations of the chosen protease assay is its inability to distinguish between different protease species within the whole

protease population. The activity of different type of proteases (e.g. acidic, serine, metallo-) could either increase or decrease under different cell culture conditions. In order to identify the individual classes of proteases and how their activity changes with changes upstream, more advanced proteomic studies should be investigated. Future work could investigate the use of alternative proteomic methods such as mass spectrometry alone, or in combination with 2D-PAGE gels or liquid chromatography (LC-MS/MS). These methods have the ability to detect and monitor multiple proteins, in the same sample. So far it has been shown that it can be evaluated how the total levels of HCPs at harvest as well as the total protease activity within these HCP populations can change when the process parameters are changing. This future work would help provide a more in-depth understanding of how the activity of individual proteases might vary between different samples and over time.

It was previously shown that increases in antibody titres can be driven by two main mechanisms: increases in cell densities or specific cell productivities, each with a different impact on resulting HCP levels (Fig. 3-3). It would be valuable to investigate if the HCP profiles as well as the levels of certain problematic HCPs differ between processes in which an increase in antibody concentration is achieved through different mechanisms. For this cell line, the relationship between the specific productivity for the antibody ( $q_{\text{Mab}}$ ) and the specific productivity for HCPs ( $q_{\text{HCP}}$ ) was evaluated. It was found that  $q_{\text{Mab}}$  and  $q_{\text{HCP}}$  are positively correlated, despite the different combinations of conditions within the DoE design (Fig. 3-4). This might suggest that an increase in  $q_{\text{Mab}}$  could lead to a difference in expression for certain problematic HCPs.

## References

- Abdi H, Williams LJ. 2013. Partial least squares methods: partial least squares correlation and partial least square regression. *Methods in molecular biology*. 930: 549-579.
- Aboulaich N, Chung WK, Thompson JH, Larkin C, Robbins D, Zhu M. 2014. A novel approach to monitor clearance of host cell proteins associated with monoclonal antibodies. *Biotechnol Prog* 30(5): 1114-1124.
- Abu-Absi SF, Yang L, Thompson P, Jiang C, Kandula S, Schilling B, and Shukla AA. 2010. Defining process design space for monoclonal antibody cell culture. *Biotechnol Bioeng* 106: 894-905.
- Adams CP, Brantner VV. 2010. Spending on new drug development. *Health Economics* 19: 130-141.
- Aho K, Derryberry D, Peterson T. 2014. Model selection for ecologists: the worldviews of AIC and BIC. *Ecology* 95(3): 631-636.
- Aldington S, Bonnerjea J. 2007. Scale-up of monoclonal antibody purification processes. *J Chrom B* 848: 64-78.
- Alexopoulos EC. 2010. Introduction to Multivariate regression analysis. *Hippokratia* 14 (Suppl 1): 23-28.
- Allmendinger R, Simaria AS, Turner R, Farid SS. 2014a. Closed-loop optimization of chromatography column sizing strategies in biopharmaceutical manufacture. *J Chem Technol Biotechnol* 89(10): 1481-1490.
- Allmendinger R, Simaria AS, Farid SS. 2014b. Multiobjective evolutionary optimization in antibody purification process design. *Biochem Eng J* 91: 250-264.
- Amanullah A, Otero JM, Mikola M, Hsu A, Zhang J, Aunins J, Schreyer HB, Hope JA, Russo AP. 2010. Novel micro-bioreactor high throughput technology for cell culture process development: Reproducibility and scalability assessment of fed-batch CHO cultures. *Biotechnol Bioeng* 106(1): 57-67.
- Anderson DR, Burnham KP, Thompson WL. 2000. Null hypothesis testing: problems, prevalence and an alternative. *Journal of Wildlife Management* 64:912-923.

- Avorn J. 2015. The \$2.6 billion pill – Methodologic and policy considerations. *N Engl J Med.* 372: 1877- 1879.
- Bailey JE. 1998. Mathematical modelling and analysis in biochemical engineering: Past Accomplishments and future opportunities. *Biotechnol Prog* 14: 8-20.
- Bailey-Kellogg C, Gutierrez AH, Moise L, Terry F, Martin WD, De Groot AS. 2014. CHOPPI: A web tool for the analysis of immunogenicity risk from host cell proteins in CHO-based protein production. *Biotechnol Bioeng* 111(11): 2170-2182.
- Banerjee A. 2010. Designing in quality: Approaches to defining the design space for a monoclonal antibody process. *BioPharm Int* 23(5)
- Baranyi J., Pin C. 2001. Modelling microbiological safety. *Food Process Modelling.* 383-401.
- Bareither R, Pollard D. 2010. A review of advanced small-scale parallel bioreactor technology for accelerated process development: Current state and future trend. *Biotechnol Prog* 27(1): 2-14.
- BCCResearch. 2012. Antibody Drugs: Technologies and Global Markets. BCCResearch.
- Becerra S, Berrios J, Osses N, Altamirano C. 2012. Exploring the effect of mild hypothermia on CHO cell productivity. *Biochem Eng J* 60: 1-8.
- Beckmann TF, Kramer O, Klausning S, Heinrich C, Thute T, Buntmeyer H, Hoffrogge R, Noll T. 2012. Effect of high passage cultivation on CHO cells: a global analysis. *Appl. Microbiol Biotechnol.*, 94: 659-671.
- Berlec A, Strukelj B. 2013. Current state and recent advances in biopharmaceutical production in *Escherichia coli*, yeasts and mammalian cells. *J Ind Microbiol Biotechnol* 40: 257-274.
- Berndt ER, Cockburn IM. 2013. Price indexes for clinical trial research: A feasibility study. National Bureau of Economic Research.
- Betts JJ, Baganz F. 2006. Miniature bioreactors: current practices and future opportunities. *Microbial Cell Factories* 5: 21.
- Biwer A, Griffith S, Cooney C. 2005. Uncertainty analysis of penicillin V production using monte carlo simulation. *Biotechnol Bioeng* 90(2): 167- 179.

Bracewell DG, Smales CM. 2013. The challenges of product- and process related impurities to an evolving biopharmaceutical industry. *Bioanalysis*, 5(2): 123-126.

Bracewell DG, Francis R, Smales CM. 2015. The future of host cell protein (HCP) identification during process development and manufacturing linked to a risk based management for their control. *Biotechnol. Bioeng.* Volume

Brodski I, Zhang Cheng, Yigzaw Y, Vedantham G. 2012. Caprylic acid precipitation method for impurity reduction: An alternative to conventional chromatography for monoclonal antibody purification. *Biotechnol Bioeng* 109(10): 2589-2598.

Broly H, Mitchell Logean C, Costioli MC, Guillemot Potelle C. 2010 Cost of Goods Modelling and Quality by Design for Developing Cost Effective Processes. *BioPharm Int* 23(6).

Büchs J, Maier U, Milbradt C, Zoels B. 2000. Power consumption in shaking flasks on rotary shaking machines: I. power consumption measurement in unbaffled flasks at low liquid viscosity. *Biotechnol Bioeng* 68: 589- 593.

Büchs J. 2001. Introduction to advantages and problems of shaken cultures. *Biochem Eng J* 7: 91-98.

Burnham KP, Anderson DR. 2002. Model selection and multi-model inference: a practical information-theoretic approach. Springer, New York, USA.

Butler M, Menesses-Acosta A. 2012. Recent advances in technology supporting biopharmaceutical production from mammalian cells. *Appl. Microbiol Biotechnol* 96: 885-894.

Capito F, Skudas R, Kolmar H, Stanislawski B. 2013. Host cell protein quantification by fourier transform mid infrared spectroscopy (FT-MIR). *Biotechnol Bioeng* 110(1): 253-258.

Chames P, Van Regenmortel M, Weiss E, Baty D. 2009. Therapeutic antibodies: successes, limitations and hopes for the future. *Br J Pharmacol* 157(2): 220-233.

Chatfield C. 2014. Durbin-Watson test. *Wiley StatsRef: Statistics reference online.*

Chen A, Chitta R, Chang D, Amanullah A. 2009. Twenty-four well plate miniature bioreactor system as a scale-down model for cell culture process development. *Biotechnol Bioeng* 102: 148-160.

Chon JH, Zarbis-Papastoitsis G. 2011. Advances in the production and downstream processing of antibodies. *New Biotechnol* 28(5): 458-463.

Chuppa S, Tsai Y-S, Yoon S, Shackelford S, Rozales C, Bhat R, Tsay G, Matanguihan C, Konstantinov K, Naveh D. 1997. Fermentor temperature as a tool for control of high-density perfusion cultures of mammalian cells. *Biotechnol Bioeng* 55:328– 338.

Clark KJR, Chaplin FWR, Harcum SW. 2004. Temperature effects on product-quality-related enzymes in batch CHO cell cultures producing recombinant tPA. *Biotechnol Prog* 20: 1888-1892.

Coco-Martin, JM, Harmsen, MM. 2008. A Review of Therapeutic Protein. *BioProcess Int* 28 - 33.

Collier R. 2009. Drug development cost estimates hard to swallow. *Canadian Medical Association* 180: 279-280.

Cordoba AJ, Shyong BJ, Breen D, Harris RJ. 2005. Non-enzymatic hinge region fragmentation of antibodies in solution. *J Chrom. B.*, 818: 115-121.

Cornell JA. 2002. Experiments with mixtures: Designs, models and the analysis of mixture data, 3<sup>rd</sup> ed., John Wiley & Sons, INC, New York.

Courtenay M. 2008. What's Ailing the Pharmaceutical Sector? Available from: <http://seekingalpha.com/article/81093-what-s-ailing-the-pharmaceutical-sector> [Accessed 23 October 2014].

De Alwis DM, Dutton RL, Scharer J, Moo-Young M. 2007. Statistical methods in media optimization for batch and fed-batch animal cell culture. *Bioprocess Biosyst Eng* 30(2): 107-113.

De Jesus MJ, Girard P, Bourgeois M, Baumgartner G, Jacko B, Amstutz H, Wurm FM. 2004. TubeSpin satellites: a fast track approach for process development with animal cells using shaking technology. *Biochem Eng J* 17: 217-223.

DiMasi J, Hansen R, Grabowski H. 2003. The price of innovation: new estimates of drug development costs. *J Health Econ* 22: 151–185.

Doneanu CE, Xenopoulos A, Fadgen K, Murphy J, Skilton SJ, Prentice H, Stapels M, Chen W. 2012. Analysis of host-cell proteins in biotherapeutic proteins by comprehensive online two-dimensional liquid chromatography/mass spectrometry. *mAbs* 4(1): 24-44.

Dorai H, Santiago A, Campbell M, Tang QM, Lewis MJ, Wang Y, Lu QZ, Wu SL, Hancock W. 2011. Characterization of the proteases involved in the N-terminal clipping of glucagon-like-peptide-1-antibody fusion proteins. *Biotechnol Prog* 27(1): 220-231.

Dorai H, Corisdeo S, Ellis D, Kinney C, Chomo M, Hawley-Nelson P, Moore G, Betenbaugh MJ, Ganguly S. 2012. Early prediction of instability of Chinese hamster ovary cell lines expressing recombinant antibodies and antibody-fusion proteins. *Biotechnol Bioeng* 109(4): 1016-1030.

Dorai H, Liu S, Yao X, Wang Y, Tekindemir U, Lewis MJ, Wu SL, Hancock W. 2013. Proteomic analysis of bioreactor cultures of an antibody expressing CHO-GS cell line that promotes high productivity. *J Proteomics Bioinform* 6(5): 98-108.

Dorai H, Ganguly S. 2014. Mammalian cell-produced therapeutic proteins: heterogeneity derived from protein degradation. *Curr Opin Biotechnol.* 30: 198-204.

Dragomirescu M, Vintila T, Vlad-Oros B, Preda G. 2008. Stabilization of microbial enzymatic preparations used in feed industry. *Zootehnie Biotehnologi* 41(1): 69-72.

Duetz WA. 2007. Microtiter plates as mini-bioreactors: miniaturization of fermentation methods. *Trends Microbiol* 15(10): 469-475.

Dziak JJ, Coffman DL, Lanza ST, Li R. 2012. Sensitivity and specificity of information criteria (technical report 12-119). The methodology center, Pennsylvania State University, PA.

Ecker DM, Jones SD, Levine HL. 2015. The therapeutic monoclonal antibody market. *mAbs*, 7(1): 9-14.

Elliott P, Hohmann A, Spanos J. 2003. Protease expression in the supernatant of Chinese hamster ovary cells grown in serum-free culture. *Biotechnol. Lett.* 22: 1949-1952.

Eon-Duval A, Broly H, Gleisner R. 2012. Quality attributes of recombinant therapeutic proteins: an assessment of impact on safety and efficacy as part of a quality by design development approach. *Biotechnol Prog* 28(3): 608-622.

Eon-Duval A, Valax P, Solacroup T, Broly H, Gleixner R, Le Strat C, Sutter J. 2012b. Application of the quality by design approach to the drug substance manufacturing process of an Fc Fusion protein: Towards a global multi-step design space. *J Pharm Sci* 101(10): 3604-3618.

Farid SS, Novais JL, Karri S, Washbrook J, Titchener-Hooker NJ. 2000 A tool for modeling strategic decisions in cell culture manufacturing. *Biotechnol Prog* 16(5): 829- 836.

Farid S. 2001. A decision-support tool for simulating the process and business perspectives of biopharmaceutical manufacture. Thesis (Ph.D.), University College London.

Farid SS, Washbrook J, Titchener-Hooker NJ. 2005a. Decision-support tool for assessing biomanufacturing strategies under uncertainty: stainless steel versus disposable equipment for clinical trial material preparation. *Biotechnol Prog*, 21(2): 486-497.

Farid SS, Washbrook J, Titchener-Hooker NJ. 2005b. Combining multiple quantitative and qualitative goals when assessing biomanufacturing strategies under uncertainty. *Biotechnol Prog* 21: 1183-1191.

Farid SS. 2007a. Process economics of industrial monoclonal antibodies manufacture. *J Chromatography B* 848: 8-18.

Farid SS, Washbrook J, Titchener-Hooker NJ. 2007b. Modelling biopharmaceutical manufacture: Design and implementation of SimBiopharma. *Computers and Chem Eng* 31: 1141-1158.

Farid S. 2009. Process economic drivers in industrial monoclonal antibody manufacture. In: *Process Scale Purification of Antibodies*. John Wiley and Sons 239-261.

Fike R. 2009. Nutrient supplementation strategies for biopharmaceutical production Part 2. Feeding for optimal recombinant protein production. *BioProcess Int* 46-52.

Fox SR, Patel UA, Yap MG, Wang DI. 2004. Maximizing Interferon-g Production by Chinese Hamster Ovary Cells by Temperature Shift Optimization: Experimental and Modeling. *Biotechnol Bioeng* 85(2): 177-184.

Frame KK, Hu WS. 1991. Kinetic study of Hybridoma cell growth in continuous culture I. A model for non-producing cells. *Biotechnol Bioeng* 37: 55-64.

Frohlich H, Villian L, Melzner D, Strube J. 2012. Membrane technology in bioprocess science. *Chem Ing Tech* 84: 905-917.

Galvanaus V, Simutis R, Lübbert A. 2004. Hybrid process models for process optimisation, monitoring and control. *Bioprocess Biosyst Eng.* 26: 393-400.

Gao SX, Zhang Y, Stansberry-Perkins K, Buko A, Bai S, Nguyen V, Brader ML. 2011. Fragmentation of a highly purified monoclonal antibody attributed to residual CHO cell protease activity. *Biotechnol Bioeng* 108(4): 977-982.

Glacken MW, Huang C, Sinskey AJ. 1989. Mathematical descriptions of Hybridoma culture kinetics III. Simulation of fed-batch bioreactor. *J Biotechnol* 10: 39-66.

George E, Titchener-Hooker NJ, Farid SS. 2007. A multi-criteria decision-making framework for the selection of strategies for acquiring biopharmaceutical manufacturing capacity. *Computers and Chemical Engineering* 31: 889-901.

Gerrodette T. 2011. Interference without significance: measuring support for hypotheses rather than rejecting them. *Marine Ecology: An evolutionary perspective* 32:404-418.

Gommers F, Perry RN, Nijhuis M, Eimers R. 2014. Monte Carlo simulations: A practical tool for setting process proven acceptable ranges for an antibody producing cell culture manufacturing process. Poster Cell Culture Engineering XIV, Quebec, Canada.

Gronemeyer P, Ditz R, Strube J. 2014. Trends in upstream and downstream process development for antibody manufacturing. *Bioengineering* 1: 188-212.

Grzeskowiak JK, Tscheliessnig A, Toh PC, Chusainow J, Lee YY, Wong N, Jungbauer A. 2009. 2-D DIGE to expedite downstream process development for human monoclonal antibody purification. *Protein Expr Purif* 66: 58-65.

Guiochon G, Beaver LA. 2011. Separation science is the key to successful biopharmaceuticals. *J Chrom A* 1218: 8836-8858.

Gutierrez AH, Moise L, De Groot AS. 2012. Of [Hamsters] and men: a new perspective on host cell proteins. *Hum. Vaccin. Immunother.* 8(9): 1172-1174.

Gyros. Available at <<http://www.gyros.com/products/products-optimized/gyrolab-xp-workstation/>>. Accessed [23 July 2015].

Haaland PD. 1989. Experimental design in biotechnology, Marcel Dekker, INC, New York

Harder A, Roels JA. 1982. Application of simple structured models in bioengineering. *Adv Biochem Eng* 21: 56-107.

Harrison RG, Todd PW, Rudge SR, Petrides D. 2003. *Bioseparations science and engineering*. New York: Oxford University Press 319-371.

Harms P, Kostov Y, French JA., Soliman M, Anjanappa M, Ram A, Rao G. 2006. Design and performance of a 24-station high throughput micro-bioreactor. *Biotechnol Bioeng* 93: 6-13.

Hassan SS, Farhan M, Mangayil R, Huttune H, Aho T. 2013. Bioprocess data mining using regularized regression and random forests. *BMC Systems Biology* 7(Suppl 1): S5.

Heath C, Kiss R. 2007. Cell culture process development: Advances in process engineering. *Biotechnol Prog* 23(1): 46-51.

Hinton PR. 2014. Multiple correlation and regression. *Statistics explained*. Taylor & Francis.

Hirai M, Frau N, Zarbis-Papastoitsis G, Kuczewski M. 2009. Hydrophobic membrane adsorbers for large-scale downstream processing. *BioPharm Int Supp*

Ho Y, Varley J, Mantalaris A. 2006. Development and analysis of a mathematical model for antibody-producing GS-NS0 cells under normal and hyperosmotic culture conditions. *Biotechnology Progress* 22: 1560-1569.

Hogwood CEM, Tait AS, Koloteva-Levine N, Bracewell DG, Smales CM. 2013. The dynamics of the CHO host cell protein profile during clarification and protein A capture in a platform antibody purification process. *Biotechnol Bioeng* 110(1): 240-251.

Hogwood CEM, Bracewell DG, Smales CM. 2014. Measurement and control of host cell proteins (HCPs) in CHO cell bioprocesses. *Curr Opin Biotechnol* 30:153-160.

Horvath B, Mun M, Laird MW. 2010. Characterization of a monoclonal antibody cell culture production process using a quality by design approach. *Mol Biotechnol* 45: 203-206.

Howell D. 2013. Multiple Regression. Statistical methods for psychology. Wadsworth Publishing. Chapter 15: 507-572.

Hsu W, Aulakh R, Traul D, Yuk I. 2012. Advanced microscale bioreactor system: a representative scale-down model for bench-top bioreactors. *Cytotechnology* 64: 667-678.

Jang JD, Barford JP. 2000. An unstructured kinetic model of macromolecular metabolism in batch and fed-batch cultures of hybridoma cells producing monoclonal antibody. *Biochemical Engineering Journal* 4: 153-168.

Jardon M, Garnier A. 2003. pH, pCO<sub>2</sub> and temperature effect on R-Adenovirus production. *Biotechnol Prog.* 19: 202-208.

Jiang Z, Droms K, Geng Z, Casnocha S, Xiao Z, Gorfien S, Jacobia SJ. 2012. Fed-batch cell culture process optimization. *Bioprocess International* 10(3): 40-45.

Jin M, Szapiel N, Zhang J, Hickey J, Ghose S. 2010. Profiling of host cell proteins by two-dimensional difference gel electrophoresis (2D-DIGE): Implications for downstream process development. *Biotechnol Bioeng* 105(2): 306-316.

Jones SD, Ransohoff TC, Castillo F, Riske FJ, Levine HL. 2015. High-throughput process development approaches for biopharmaceuticals. *American Pharmaceutical Review*.

Kadane JB, Lazar NA. 2004. Methods and criteria for model selection. *Journal of the American Statistical Association.* 99: 279-290.

Kaitin KI. 2010. The landscape for pharmaceutical innovation: Drivers of cost-effective clinical research. *Pharm Outsourcing* 1-6.

Kao YH, Hewitt DP, Trexler-Schmidt M, Laird MW. 2010. Mechanism of antibody reduction in cell culture production processes. *Biotechnol Bioeng* 107(4): 622-632.

Kaufmann H, Mazur X, Fussenegger M, Bailey JE. 1999. Influence of low temperature on productivity, proteome and protein phosphorylation of CHO cells. *Biotechnol Bioeng* 63(5): 573-582.

Kelley B. 2007. Very large scale monoclonal antibody purification: The case for conventional unit operations. *Biotechnol Prog* 23: 995-1008.

Kelley B. 2009. Industrialization of mAb production technology - the bioprocessing industry at a crossroads. *MAbs* 1(5): 443-452.

- Khan KH. 2013. Gene expression in mammalian cells and its applications. *Adv Pharm Bull* 3(2): 257-263.
- Kim MS, Kim NS, Sung YH, Lee GM. 2002. Biphasic culture strategy based on hyperosmotic pressure for improved humanized antibody production in chinese hamster ovary cell culture. *In Vitro Cellular & Developmental Biology – Animal*, 38: 314-319.
- Kim BJ, Zhao T, Young L, Zhou P, Shuler ML. 2011. Batch, fed-batch and microcarrier cultures with CHO cell lines in a pressure-cycle driven miniaturized bioreactor. *Biotechnol Bioeng* 109(1): 137-145.
- Kim BJ, Diao J, Shuler ML. 2012. Mini-scale bioprocessing systems for highly parallel animal cell cultures. *Biotechnol Prog* 28(3): 595-607.
- Kirdar AO, Green KD, Rathore AS. 2008. Application of multivariate data analysis for identification and successful resolution of a root cause for a bioprocessing application. *Biotechnol Prog* 24: 720-726.
- Koharyova M, Kolarova M. 2008. Oxidative stress and thioredoxin system. *Gen Physiol Biophys* 27(2): 71-84.
- Kontoravdi C, Asprey SP, Pistikopoulos EN, Mantalaris A. 2005. Application of global sensitivity analysis to determine goals for design of experiments: An example study on antibody-producing cell cultures. *Biotechnol Prog* 21: 1128-1135.
- Kompala DS. 2013. Cell growth and protein expression kinetics. *Upstream Industrial Biotechnology 2*, Chapter 6, Wiley
- Kontoravdi C, Asprey SP, Pistikopoulos EN, Mantalaris A. 2007. Development of a dynamic model of monoclonal antibody production and glycosylation for product quality monitoring. *Computers and Chemical Engineering* 31: 392-400.
- Koterba KL, Borgschulte T, Laird MW. 2012. Thioredoxin 1 is responsible for antibody disulfide reduction in CHO cell culture. *J Biotechnol* 157: 261-267.
- Kourti T. 2004. Process analytical technology and multivariate statistical control, Part 1. *Process Anal. Technol.* 1:13-19.
- Kourti T. 2010. Pharmaceutical manufacturing: the role of multivariate analysis in design space. *Control strategy, process understanding, troubleshooting, optimization.*

In David J et al (ed) Chem engineering in the pharmaceutical industry: R&D to manufacturing. Wiley, 853-878.

Kourti T, Davis B. 2012. The business benefits of Quality by Design (QbD). *Pharmaceutical Engineering* 32(4): 1-10.

Kourti T. 2015. Multivariate analysis for process understanding, monitoring, control and optimization of lyophilization processes. In Feroz J et al (ed) *Quality by design for biopharmaceutical drug product development*. Springer, 537-565.

Kumar S, Wittmann C, Heinzle E. 2004. Minibioreactors. *Biotechnol Lett* 26: 1-10.

Lain B, Zarbis-Papastoitsis G, Schirmer E, Kuczewski M. 2010. PEG Precipitation: A powerful tool for monoclonal antibody purification. *BioPharm Int Supp*

Lang HJ. 1948. Simplified approach to preliminary cost estimates. *Chemical Eng* 55: 112- 113.

Lee JM. 2001. *Biochemical Engineering*. Department of Chemical Engineering, Washington State University, Pullman, WA 99164-2710.

Lee MS, Kim KW, Kim YH, Lee GM. 2003. Proteome analysis of antibody-expressing CHO cells in response to hyperosmotic pressure. *Biotechnol Prog* 19: 1734-1741.

Lee KM, Gilmore DF. 2006. Statistical experimental design and bioprocess modeling and optimization analysis. Repeated-measures method for dynamic biotechnology process. *App Biochem and Biotechnol* 135: 101-115.

Legmann R, Schreyer HB, Combs RG, McCormick EL, Russo AP, Rodgers ST. 2009. A predictive high-throughput scale-down model of monoclonal antibody production in CHO cells. *Biotechnol Bioeng* 104(6): 1107-1120.

Levy NF, Valente KN, Choe LH, Lee KH, Lenhoff AM. 2014. Identification and characterization of host cell protein-associated impurities in monoclonal antibody bioprocessing. *Biotechnol Bioeng* 111(5): 904-912.

Lewis G, Lugg R, Wales R. 2010. Novel automated micro-scale bioreactor technology: A qualitative and quantitative mimic for early process development. *Bioprocessing Journal* 9(1): 22-25.

Li F, Zhou JX, Yang X, Tressel T, Lee B. 2005. Current therapeutic antibody production and process optimization. *BioProcessing J* Sept/Oct 1-8.

Li F, Vijayasankaran N, Shen A, Kiss R, Amanullah A. 2010. Cell culture processes for monoclonal antibody production. *Landes Bioscience Journals MAbs* 2: 466–479.

Lim AC, Zhou Y, Washbrook J, Sinclair A, Fish B, Francis R, Titchener-Hooker NJ, Farid SS. 2005. Application of a decision-support tool to assess pooling strategies in perfusion culture processes under uncertainty. *Biotechnol Prog* 21(4): 1231-1242.

Lim AC, Washbrook J, Titchener-Hooker NJ, Farid SS. 2006. A computer-aided approach to compare the production economics of fed-batch and perfusion culture under uncertainty. *Biotechnol Bioeng* 93(4): 687-697.

Lim JAC, Patkar A, McDonagh G, Sinclair A, Lucy P. 2010. Modeling bioprocess cost. Process economic benefits of expression technology based on *Pseudomonas fluorescens*. *BioProcess Int* November 62-70.

Lin J, Takagi M, Qu Y, Gao P, Yoshida T. 1999. Enhanced monoclonal antibody production by gradual increase of osmotic pressure. *Cytotechnology* 29: 27-33.

Link T, Backstrom M, Graham R, Essers R, Zorner K, Gatgens J, Burchell J, Taylor-Papadimitriou J, Hansson GC, Noll T. 2004. Bioprocess development for the production of a recombinant MUC1 fusion protein expressed by CHO-K1 cells in protein-free medium. *Journal of Biotechnology* (110): 51-62.

Liu HF, Ma J, Wnter C, Bayer R. 2010. Recovery and purification process development for monoclonal antibody production. *MAbs* 2(5): 480-499.

Looby M, Ibarra N, Pierce JJ, Buckley K, O'Donovan E, Heenan M, Moran E. 2011. Application of a Quality by Design principles to the development and technology transfer of a major process improvement for the manufacture of a recombinant protein. *Biotechnol Prog* 27(6): 1718-1729.

Lotter S, Buchs J. 2004. Utilization of specific power input measurements for optimization of culture conditions in shaking flasks. *Biochem Eng J* 17: 195-203.

Lowe CR. 2001. Combinational approaches to affinity chromatography. *Current Opinion in Chemical Biology* 5(3): 248-256.

Lu F., Toh PC, Burnett I, Li F, Hudson T, Amanullah A, Li J. 2013. Automated dynamic fed-batch process and media optimization for high productivity cell culture process development. *Biotechnol Bioeng* 110(1): 191-205.

- Lye GJ, Ayazi-Shamlou P, Baganz F, Dalby P, Woodley J. 2003. Accelerated design of bioconversion processes using automated microscale processing techniques. *Trends in Biotechnol* 21: 29-37.
- Mahajan E, George A, Wolk B. 2012. Improving affinity chromatography resin efficiency using semi-continuous chromatography. *J Chrom A* 1227: 154-162.
- Marill KA. 2004. Advanced statistics: linear regression, part II: multiple linear regression. *Academic Emergency Medicine* 11(1): 94-102.
- Mason M, Sweeney B, Cain K, Stephens P, Sharfstein ST. 2014. Reduced culture temperature differentially affects expression and biophysical properties of monoclonal antibody variants. *Antibodies* 3: 253-271.
- Mayer-Bartschmid A, Clarkson M, Zoro B, Groth M, Schubel A. 2013. New improved automation of the AMBR™ microbioreactor implemented in a clone selection workflow. Poster 23<sup>rd</sup> ESACT Meeting, June
- McDonald JH. 2008. Multiple regression. *Handbook of Biological Statistics*. Sparky House Publishing 217-223.
- Montgomery DC. 2004. *Design and analysis of experiments*, 6<sup>th</sup> ed., John Wiley & Sons, INC, New York.
- Montgomery DC. 2009. *Design and analysis of experiments*. 7<sup>th</sup> ed. New Jersey: Wiley and sons, Inc., p. 656.
- Moore A, Mercer J, Dutina G, Donahue CJ, Bauer KD, Mather JP, Etcheverry T, Ryll T. 1997. Effects of temperature shift on cell cycle, apoptosis and nucleotide pools in CHO cell batch cultures. *Cytotechnology* 23: 47-54.
- Moses S, Manahan M, Ambrogelly A, Ling WLW. 2012. Assessment of AMBR™ as a model for high- throughput cell culture process development strategy. *Adv in Bioscience and Biotechnol* 3: 918-927.
- Muller-Spath T, Aumann L, Strohle G, Kornmann H, Valax P, Delegrange L, Charbaut E, Baer G, Lamprove A, Johnck M, Schulte M, Morbidelli M. 2011. Two step capture and purification of IgG2 using multicolumn counter-current solvent gradient purification (MCSGP). *Biotechnol Bioeng* 107(6): 974-984.
- Murtaugh PA. 2014. In defence of P values. *Ecology* 95(3): 611-617.

Mustafa MA, Washbrook J, Lim AC, Zhou Y, Titchener-Hooker NJ, Morton P, Berezenko S, Farid SS. 2004. A software tool to assist business-process decision-making in the biopharmaceutical industry. *Biotechnol Prog* 20(4): 1096-1102 [Correction: 2005 *Biotechnol Prog* 21(1): 320-320].

Mustafa MA, Washbrook J, Titchener-Hooker NJ, Farid SS. 2006. Retrofit decisions within the biopharmaceutical industry. An EBA case study. *Food Bioproducts Process* 84(1): 84- 89.

Myers RH, Montgomery DC. 2002. *Response surface methodology: Process and product optimization using designed experiments*, 2nd ed.. John Wiley & Sons, INC., New York.

Nienow AW, Rielly CD, Brosnan K, Barg K, Lee K, Coopman K, Hewitt CJ. 2013. The physical characterisation of a microscale parallel bioreactor platform with an industrial CHO cell line expressing an IgG4. *Biochem Eng J* 76: 25-36.

Nogal B, Chhiba K, Emery JC. 2012. Select host cell proteins coelute with monoclonal antibodies in protein a chromatography. *Biotechnol Prog* 28(2): 454-458.

Novais JL, Titchener-Hooker NJ, Hoare M. 2001. Economic comparison between conventional and disposables-based technology for the production of biopharmaceuticals. *Biotechnol Bioeng* 75(2): 143- 153.

Oh SKW, Kuek KH, Wong VVT. 2004. Design, simulation and optimization of a large-scale monoclonal antibody production plant: Upstream design. *Pharmaceutical Eng* 24(6): 42-46.

O'Kelley M, Berry S, Walp D, Garrett A. 2012. Lifecycle modelling and simulation. [www.quintiles.com](http://www.quintiles.com)

Palmer PB, O'Connell DG. 2009. Regression analysis for prediction: Understanding the process. *Cardiopulmonary physical therapy journal* 20(3): 23-26.

Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, Schacht AL. 2010. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nature Reviews Drug Discovery* 9: 203-214.

Pezzini J, Joucla G, Gantier R, Touelle M, Lomenech AM, Le Senechal C, Garbay B, Santarelli X, Cabanne C. 2011. Antibody capture by mixed-mode chromatography: A comprehensive study from determination of optimal purification conditions to identification of contaminating host cell proteins. *J Chrom. A.* 1218: 8197-8208.

Platts K., Probert DR, Canez L. 2002. Make vs. buy decisions: A process incorporating multi-attribute decision-making. *International Journal of Production Economics* 77: 247-257.

Pollock J, Ho SV, Farid SS. 2013. Fed-batch and perfusion culture processes: Economic, environmental and operational feasibility under uncertainty. *Biotechnol Bioeng* 110(1): 206-219.

PR Newswire. 2015. Outlook of the global pharmaceutical outsourcing market 2015-2020 – Current market accounts for approx \$130.65 billion. Research and Markets. Available at < <http://www.prnewswire.co.uk/news-releases/outlook-of-the-global-pharmaceutical-outsourcing-market-2015-2020---current-market-accounts-for-approx-13065-billion-499341031.html>> Accessed [23 June 2015]

Provost A, Bastin G. 2004. Dynamic metabolic modelling under the balanced growth condition. *J. Process Control* 14: 717-728.

Puich M, Paz A. 2004. Simulations improve production capacity. *BioPharm Int*, May 1.

Qureshi AS, Bhutto MA, Khushk I, Dahot MU. 2011. Optimization of cultural conditions for protease production by *Bacillus subtilis* EFRL 01. *African J Biotechnol* 10(26): 5173-5181.

Rader RA. 2013. FDA Biopharmaceutical Product approvals and trends in 2012. *Bioprocess International* 11(3): 18-27.

Rameez S, Mostafa SS, Miller C, Shukla AA. 2014. High-throughput miniaturized bioreactors for cell culture process development: reproducibility, scalability and control. *Biotechnol Prog* 30(3): 718-727.

Ransohoff TC. 2009. If you build it, will they come? The promise and perils of investing in biomanufacturing capacity. *Bioprocess Technology Consultants, Inc.*, 2<sup>nd</sup> Annual Sanford C. Bernstein Biosimilars Conference, New York, November 19.

- Rao RV. 2007. Introduction to multiple attribute decision-making (MADM) methods. Decision making in manufacturing environment using graph theory and fuzzy multiple attribute decision making methods. 2: 27-41.
- Rathore AS, Latham P, Kaltenbrunner O, Curling J, Levine H. 2004. Costing issues in the production of biopharmaceuticals. *BioPharm Int*, February.
- Rathore AS, Johnson R, Yu O, Kirdar AO, Annamalai A, Ahuja S, Ram K. 2007. Applications of multivariate data analysis in biotech processing. *BioPharm Int*
- Rathore AS, Bhushan N, Hadpe S. 2011. Chemometrics application in biotech processes: review. *Biotechnol Prog* 27: 307-315.
- Rathore AS, Mittal S, Pathak M, Arora A. 2014. Guidance for performing multivariate data analysis of bioprocessing data: Pitfalls and recommendations. *Biotechnol Prog*. 30(4): 967-973.
- Ratner B. 2010. Variable selection methods in regression: Ignorable problem, outing notable solution. *Journal of Targeting, Measurement and Analysis for Marketing* 18: 65-75.
- Razali NM, Wah YB. 2011. Power comparison of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of statistical modelling and analytics* 2(1): 21-33.
- Reichenberger G. 2011. Improving immunoassay performance. New tools aim to reduce turnaround times in biotherapeutic development and production. [Online]. Available at: <<http://gen.epubxp.com/i/64286/30>> Accessed [23 July 2014]
- Reisinger V, Toll H, Mayer RE, Visser J, Wolschin F. 2014. A mass spectrometry-based approach to host cell protein identification and its application in a comparability exercise. *Analytical Biochem* 463: 1-6.
- Remer DS, Idrovo JH. 1991. Cost-estimating factors for biopharmaceutical process equipment. *Int J Pharmacy and Technology* 36-42.
- Riley BS, Li X. 2011. Quality by design and process analytical technology for sterile products –Where are we now? *AAPS Pharm Sci Tech* 12(1): 114-118.
- Robert F, Bierau H, Rossi M, Agugiaro D, Soranzo T, Broly H, Mitchell-Logean C. 2009. Degradation of an Fc-fusion recombinant protein by host cell proteases: Identification of a CHO cathepsin D protease. *Biotechnol Bioeng* 104(6): 1132-1141.

Rodrigues ME, Costa AR, Henriques M, Azeredo J, Oliveira R. 2010. Technological progresses in monoclonal antibody production systems. *Biotechnol Prog* 26(2): 332-351.

Roubos JA, Babuska R, Krabben P, Heijnen JJ. 2000. Hybrid modeling of fed-batch bioprocesses; combination of physical equations with metabolic networks and black-box kinetics *Journal A, Benelux Q J Automatic Control* 41:17–23.

Roubos H. 2002. Bioprocess modelling and optimisation. PhD thesis, Delft University of Technology, The Netherlands.

Rouiller Y, Solacroup T, Deparis V, Barbafieri M, Gleixner R, Broly H, Eon-Duval A. 2012. Application of quality by design to the characterization of the cell culture process of an Fc-Fusion protein. *European journal of pharmaceutics and biopharmaceutics* 81(2): 426-437.

Roy J. 2011. Drugs, medicines and regulatory authorities. Chapter 3. An introduction to *Pharmaceutical Sciences: Production, Chemistry, Techniques and Technology*. 43-69.

Ryll T, Dutina G, Reyes A, Gunson J, Krummen L, Etcheverry T. 2000. Performance of small-scale perfusion cultures using an acoustic cell filtration device for cell retention: characterization of separation efficiency and impact of perfusion on product quality. *Biotechnol Bioeng* 69: 440– 449.

Ryu JS, Kim TK, Chung JY, Lee GM. 2000. Osmoprotective effect of glycine betaine on foreign protein production in hyperosmotic recombinant chinese hamster ovary cell cultures differs among cell lines. *Biotechnol Bioeng* 70(2): 168-175.

Sandberg H, Lutkemeyer D, Kuprin S, Wrangel M, Almstedt A, Persson P, Ek V, Milkaelsson M. 2006. Mapping and partial characterization of proteases expressed by a CHO production cell line. *Biotechnol Bioeng* 95(5): 961-971.

Saraswat M, Musante L, Ravida A, Shortt B, Byrne B, Holthofer H. 2013. Preparative purification of recombinant proteins: Current status and future trends. *BioMed Research International* 1-18.

Scott C. 2012. A Decade of Product Development. *BioProcess International*, 10, 72-78.

Schirmer EB, Kuczewski M, Golden K, Lain B, Bragg C, Chon L, Cacciuttolo M, Zarbis-Papastoitsis. 2010. Primary clarification of very high-density cell culture harvests by enhanced cell settling. *BioProcess Int* 32-29.

Schnatz RG. 2013. Modern-day drug discovery and development. Remington-Essentials of Pharmaceutics. Chapter 7. Pharmaceutical Press 81-93.

Schubert J, Simutis R, Dors M, Havlik I, Lubbert A. 1994. Bioprocess optimization and control: application of hybrid modelling. *J Biotechnol* 35: 51–68.

Schwartz WR, Kembhavi A, Harwood D, Davis LS. 2009. Human detection using partial least squares analysis. *Int. Conf. on Computer vision and pattern recognition (CVPR)*.

Sekhon BS. 2010. Biopharmaceuticals: an overview. *Thai J Pharm Sci* 34: 1-19.

Seo JS, Kim YJ, Cho JM, Baek E, Lee GM. 2013. Effect of culture pH on recombinant antibody production by a new human cell line, F2N78, grown in suspension at 33.0 °C and 37.0 °C. *Appl. Microbiol Biotechnol* 97(12): 5283- 5291.

Shevitz J, Bonham Carter J, Lim J, Sinclair A. 2011. An economic comparison of three cell culture techniques. *Biopharm Int* 24(2).

Shukla AA, Hubbard B, Tressel T, Gunhan S, Low D. 2007. Downstream processing of monoclonal antibodies-application of platform approaches. *J Chromat B* 848: 29-39.

Shukla AA, Hinckley P. 2008. Host cell protein clearance during protein A chromatography: development of an improved column wash step. *Biotechnol Prog* 24: 1115-1121.

Shukla AA, Jiang C, Ma J, Rubacha M, Flansburg L, Lee SS. 2008. Demonstration of robust host cell protein clearance in biopharmaceutical downstream processes. *Biotechnol Prog* 24: 614-622.

Shukla AA, Thömmes J. 2010. Recent advances in large-scale production of monoclonal antibodies and related proteins. *Trends Biotechnol* 28(5): 253–261.

Sidoli FR, Mantalaris A, Asprey SP. 2004. Modelling of mammalian cells and cell culture processes. *Cytotechnology* 44: 27-46.

Sidoli FR, Mantalaris A, Asprey SP. 2005. Toward global parametric estimability of a large-scale kinetic single-cell model for mammalian cell cultures. *Industrial & engineering chemistry research* 44: 868-878.

Silk NJ, Denby S, Lewis G, Kuiper M, Hatton D, Field R, Baganz F, Lye GJ. 2010. Fed-batch operation of an industrial cell culture process in shaken microwells. *Biotechnol Lett* 32(1): 73-78.

Simaria AS, Turner R, Farid SS. 2012. A multi-level meta-heuristic algorithm for the optimisation of antibody purification processes. *Biochem Eng J* 69: 144-154.

Sinclair A. 2010. How geography affects the cost of biomanufacturing. *BioProcess Int* June 1.

Sinnott, RK. 1993. *Coulson and Richardson's Chemical Engineering*. Oxford: Pergamon Press 6: 209-244.

Sisodiya VN, Lequieu J, Rodriguez M, McDonald P, Lazzareschi KP. 2012. Studying host cell protein interactions with monoclonal antibodies using high throughput protein A chromatography. *Biotechnol. J.* 7(10): 1233- 1241.

Sommerfeld S, Strube J. 2005. Challenges in biotechnology production- generic processes and process optimization for monoclonal antibodies. *Chem Eng Proc* 44: 1123-1137.

Steuer RE, Na P. 2003. Multiple criteria decision making combined with finance: A categorized bibliographic study. *European Journal of Operational Research* 150(3): 496-515.

Stockdale GW, Cheng A. 2009. Design Space and reliable operating region using a multivariate Bayesian approach with experimental design. *Quality Technol & Quantitative Management* 6(4): 391-408.

Stonier A, Simaria AS, Smith M, Farid SS. 2012. Decisional tool to assess current and future process robustness in an antibody purification facility. *Biotechnol Prog* 28(4): 1019-1028.

Stonier A, Pain D, Westlake A, Hutchinson N, Thornhill NF, Farid SS. 2013. Integration of stochastic simulation with multivariate analysis: Short-term facility fit prediction. *Biotechnol Prog* 29(2): 368-377.

Tait AS, Hogwood CEM, Smales CM, Bracewell DG. 2012. Host cell protein dynamics in the supernatant of a mAb producing CHO cell line. *Biotechnol Bioeng* 109(4): 971-982.

Tait AS, Tarrant RD, Velez-Suberbie ML, Spencer DL, Bracewell DG. 2013. Differential response in downstream processing of CHO cells grown under mild hypothermic conditions. *Biotechnol Prog* 29(3): 688-696.

Tao Y, Ibraheem A, Conley L, Cecchini D, Ghose S. 2014. Evaluation of high-capacity cation exchange chromatography for direct capture of monoclonal antibodies from high-titer cell culture processes. *Biotechnol Bioeng* 111(7): 1354-1364.

Tarpey T. 2000. A note on the prediction sum of squares statistic for restricted least squares. *The American Statistician*. Taylor & Francis 54(2): 116-118.

Tarrant RDR, Velez-Suberbie ML, Tait AS, Smales CM, Bracewell DG. 2012. Host cell protein adsorption characteristics during Protein A chromatography. *Biotechnol Prog*. 28(4): 1037-1044.

Tcheliessnig AL, Konrath J, Bates R, Jungbauer A. 2013. Host cell protein analysis in therapeutic protein bioprocessing – methods and applications. *Biotechnol J* 8: 655-670.

Thabane L, Mbuagbaw L, Zhang S, Samaan Z, Marcucci M, Ye C, Thabane M, Giangregorio L, Dennis B, Kosa D, Borg Debono V, Dillenburg R, Fruci V, Bawor M, Lee J, Wells G, Goldsmith CH. 2013. A tutorial on sensitivity analyses in clinical trials: the what, why, when and how. *BMC Medical Research Methodology* 13:92.

Thompson JH, Chung WK, Zhu M, Tie L, Lu Y, Aboulaich N, Strouse R, Mo W. 2014. Improved detection of host cell proteins (HCPs) in a mammalian cell-derived antibody drug using liquid chromatography/mass spectrometry in conjunction with an HCP-enrichment strategy. *Rapid Commun Mass Spectrom* 28: 855-860.

Touelle M, Uzel A, Depoisier JF, Gantier R. 2011. Designing new monoclonal antibody purification processes using mixed-mode chromatography sorbents. *J Chrom B* 879(13-14): 836-843.

Trexler-Schmidt M, Sargis S, Chiu J, Sze-Khoo S, Mun M, Kao Y-H, Laird MW. 2010. Identification and prevention of antibody disulfide bond reduction during cell culture manufacturing. *Biotech Bioeng* 106(3): 452-461.

Triantaphyllou E. 2000. Multi-criteria decision making methods: A comparative study. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Trummer E, Fauland K, Seidinger S, Schriebl K, Lattenmayer C, Kunert R, Vorauer-Uhl K, Weik R, Borth N, Katinger H, Muller D. 2006. Process parameter shifting: Part I. Effect of DOT, pH and temperature on the performance of Epo-Fc expressing CHO cells cultivated in controlled batch bioreactors. *Biotechnol Bioeng* 94(6): 1033-1044.

Tsao YS, Cardoso AG, Condon RGG, Voloch M, Lio P, Lagos JC, Kearns BG, Liu Z. 2005. Monitoring chinese hamster ovary cell culture by the analysis of glucose and lactate metabolism. *Journal of Biotechnology* 118: 316-327.

Tsuchiya HM, Fredrickson AG, Aris R. 1966. Dynamics of Microbial Cell Populations. *Adv. Chem. Eng.* 6: 125-206.

Van Amum P. 2015. Big Pharma up biologics manufacturing. Available at: <<http://connect.dcat.org/blogs/patricia-van-arnum/2015/03/03/big-pharmas-ramps-up-in-biologics-manufacturing>> Accessed [20 June 2015]

Vanderlaan M, Sandoval W, Liu P, Nishihara J, Tsui G, Lin M, Parker S, Wong RM, Low J, Wang X, Yang J, Veeravalli K, McKay P, O'Connell L, Tran B, Vij R, Fong C, Francissen F, Zhu-Simoni J, Quarmby V, Krawitz D. 2015. Hamster Phospholipase B-like 2 (PLBL2), a host cell protein impurity in CHO-derived therapeutic monoclonal antibodies. *BioProcess Int.* April.

Valente KN, Lenhoff AM, Lee KH. 2015. Expression of difficult-to-remove host cell protein impurities during extended chinese hamster ovary cell culture and their impact on continuous bioprocessing. *Biotechnol Bioeng* 112(6): 1232-1242.

Vrieze SI. 2012. Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychol Methods* 17(2): 228-243.

Wang MD, Yang M, Huzel N, Butler M. 2002. Erythropoietin production from CHO cells grown in continuous culture in a fluidized-bed bioreactor. *Biotechnol Bioeng* 77: 194-203.

Wang X, Hunter AK, Mozier NM. 2009. Host cell proteins in biologics development: Identification, quantification, and risk assessment. *Biotechnol Bioeng* 103(3): 446-458.

- Wang H. 2010. Stochastic modeling of the equilibrium speed-density relationship. Thesis (Ph.D.), University of Massachusetts, USA. Available from: <http://www.scribd.com/doc/59861228/4/Deterministic-vs-stochastic-models> [Accessed 12 May 2015].
- Wang W, Ignatius AA, Thakkar SV. 2014. Impact of residual impurities and contaminants on protein stability. *J Pharma Sci.* 103(5): 1315-1330.
- Werner RG. 2004. Economic aspects of commercial manufacture of biopharmaceuticals. *J Biotechnol* 113: 171-182.
- Westoby M, Chrostowski J, de Vilmorin P, Smelko JP, Romero JK. 2011. Effects of solution environment on mammalian cell fermentation broth properties: enhanced impurity removal and clarification performance. *Biotechnol Bioeng* 108(1): 50-8.
- Williams MN, Grajales CAG, Kurkiewicz D. 2013. Assumptions of multiple regression: Correcting two misconceptions. *Practical Assessment, Research & Evaluation* 18(11): 1-14.
- Wittmann C, Kim HM, Heinzle E, John G. 2003. Characterization and application of an optical sensor for quantification of dissolved O<sub>2</sub> in shake-flasks. *Biotechnol Lett*, 25: 377-380.
- Wlaschin KF, Hu WS. 2006. Fed-batch culture and dynamic nutrient feeding. *Adv.Biochem. Eng. Biotechnol.* 101: 43-74.
- Wu P. 2013. Practical experiences integrating upstream and downstream processing. *Am Pharm Rev* 16: 84-86.
- Wurm FM. 2015. CHO History, CHO Evolution and CHO Genomics – an Unsolvable Enigma? *Animal Cell Biotechnology: In Biologics Production.* De Gruyter. 38-59.
- Xie L, Wang DIC. 1994. Fed-batch cultivation of animal cells using different medium design concepts and feeding strategies. *Biotechnol Bioeng* 43: 1175-1189.
- Yang Y, Farid SS, Thornhil NF. 2014. Data mining for rapid prediction of facility fit and debottlenecking of biomanufacturing facilities. *J Biotechnol* 179: 17-25.
- Yeatts AB, Fisher JP. 2011. Bone tissue engineering bioreactors: Dynamic culture and the influence of shear stress. *Bone* 48(2): 171-181.

Yigzaw Y, Piper R, Tran M, Shukla AA. 2006. Exploitation of the adsorptive properties of depth filters for host cell protein removal during monoclonal antibody purification. *Biotechnol. Prog* 22: 288-296.

Yoon SK, Song JY, Lee GM. 2003. Effect of low temperature on specific productivity, transcription level and heterogeneity of erythropoietin in Chinese hamster ovary cells. *Biotechnol Bioeng* 82(3): 289- 298.

Yoon SK, Choi SL, Song JY, Lee GM. 2005. Effect of culture pH on erythropoietin production by Chinese hamster ovary cells grown in suspension at 32.5 and 37 °C. *Biotechnol Bioeng* 89(3): 345- 356.

Yu M, Hu Z, Pacis E, Vijayasankaran N, Shen A, Li F. 2011. Understanding the intracellular effect of enhanced nutrient feeding towards high titer antibody production process. *Biotechnol Bioeng* 108(5): 1078-1088.

Yuk IN, Nishihara J, Walker Jr D, Huang E, Gunawan F, Subramanian J, Pynn AFJ, Yu XC, Zhu-Simoni J, Vanderlaan M, Krawitz DC. 2015. More similar than different: Host cell protein production using three null CHO cell lines. *Biotechnol Bioeng* finish

Zeng AP. 1996. Mathematical modelling and analysis of monoclonal antibody production by Hybridoma cells. *Biotechnol Bioeng* 50: 238-247

Zhang X, Chang Z. 2004. Temperature dependent protease activity and structural properties of human HtrA2 protease. *Biochemistry (Moscow)* 69(6): 687-692.

Zhang J. 2011. New global pharmaceutical outsourcing trends. *Pharmaceutical online*. Available at <<http://www.pharmaceuticalonline.com/doc/new-global-pharmaceutical-outsourcing-trends-0001>> Accessed [23 June 2015]

Zhang Q, Goetze AM, Cui H, Wylie J, Trimble S, Hewig A, Flynn GC. 2014. Comprehensive tracking of host cell proteins during monoclonal antibody purification using mass spectrometry. *mAbs* 6(3): 659-670.

Zhu J. 2012. Mammalian cell protein expression for biopharmaceutical production. *Biotechnol Adv* 30(5): 1158-1170.

Zhu-Simoni J, Yu C, Nishihara J, Wong RM, Gunawan F, Lin M, Krawitz D, Liu P, Sandoval W, Vanderlaan M. Host cell protein testing by ELISAs and the use of orthogonal methods. *Biotechnol Bioeng* 111(12): 2367-2379.

## **Appendix**

### **Chapter 5 Appendix**

**Table 5-1A: Summary statistics for eight “best” models for AMBR 2 titre presented in coded factors. The coded factors B, D and E refer to the cell culture variables seeding density, pH and feed start time, respectively**

	AMBR 2 – Titre	STATISTICS					
Model terms	Model	R <sup>2</sup>	Predicted R <sup>2</sup>	Adjusted R <sup>2</sup>	R <sup>2</sup> k fold	RMSE	AIC
1	D*	0.58	0.50	0.56	0.49	1.307	86.09
2	D*+D <sup>2</sup> *	0.96	0.94	0.95	0.94	0.430	34.55
3	D*+E+D <sup>2</sup> *	0.96	0.94	0.96	0.94	0.401	34.23
4	D*+E*+DE*+D <sup>2</sup> *	0.97	0.95	0.97	0.95	0.346	28.50
5	B+D*+E+DE*+D <sup>2</sup> *	0.98	0.95	0.97	0.95	0.345	31.12
<b>6</b>	<b>B+D*+E*+BD*+DE*+D<sup>2</sup>*</b>	<b>0.98</b>	<b>0.97</b>	<b>0.98</b>	<b>0.97</b>	<b>0.289</b>	<b>25.65</b>
7	B+D*+E*+BD*+BE+DE*+D <sup>2</sup> *	0.99	0.94	0.98	0.94	0.268	26.12
8	B+D*+E*+BD*+BE+DE*+B <sup>2</sup> +D <sup>2</sup> *	0.99	0.94	0.98	0.94	0.272	31.23
9	B+D*+E*+BD*+BE+DE*+B <sup>2</sup> +D <sup>2</sup> *+E <sup>2</sup> *	0.99	0.92	0.98	0.93	0.281	38.24

**Table 5-2A: Summary statistics for eight “best” models for AMBR 2 HCP presented in coded factors. The coded factors B, D and E refer to the cell culture variables seeding density, pH and feed start time, respectively.**

	AMBR 2 – HCP	STATISTICS					
Model terms	Model	R <sup>2</sup>	Predicted R <sup>2</sup>	Adjusted R <sup>2</sup>	R <sup>2</sup> k fold	RMSE	AIC
1	D*	0.77	0.73	0.76	0.74	369805	688.62
2	D*+D <sup>2</sup> *	0.95	0.94	0.94	0.94	177531	655.18
3	D*+E+D <sup>2</sup> *	0.96	0.94	0.95	0.94	161510	652.70
4	B*+D*+B <sup>2</sup> *+D <sup>2</sup> *	0.97	0.94	0.96	0.94	153935	652.77
5	B+D*+E+B <sup>2</sup> *+D <sup>2</sup> *	0.97	0.94	0.96	0.94	142699	651.90
6	B*+D*+E+DE+B <sup>2</sup> *+D <sup>2</sup> *	0.98	0.94	0.97	0.94	136509	652.99
7	<b>B+D*+E+BE+B<sup>2</sup>*+D<sup>2</sup>*+E<sup>2</sup>*</b>	<b>0.98</b>	<b>0.95</b>	<b>0.97</b>	<b>0.95</b>	<b>125536</b>	<b>652.77</b>
8	B+D*+E+BD+BE*+B <sup>2</sup> *+D <sup>2</sup> *+E <sup>2</sup> *	0.98	0.94	0.97	0.93	127115	657.89
9	B+D*+E+BD+BE*+DE+B <sup>2</sup> *+D <sup>2</sup> *+E <sup>2</sup> *	0.98	0.94	0.97	0.92	130201	664.46

**Table 5-3A: Summary statistics for eight “best” models for AMBR 3 titre presented in coded factors. The coded factors A, B and C refer to the cell culture variables media osmolality, seeding density and temperature, respectively.**

	AMBR 3 – Titre	STATISTICS					
Model terms	Model	R <sup>2</sup>	Predicted R <sup>2</sup>	Adjusted R <sup>2</sup>	R <sup>2</sup> k fold	RMSE	AIC
1	C*	0.70	0.64	0.68	0.65	0.605	47.30
2	B*+C*	0.82	0.74	0.80	0.73	0.481	38.60
3	B*+C*+BC*	0.86	0.77	0.83	0.76	0.436	36.21
<b>4</b>	<b>A+B*+C*+A<sup>2</sup>*</b>	<b>0.91</b>	<b>0.82</b>	<b>0.89</b>	<b>0.82</b>	<b>0.351</b>	<b>28.68</b>
5	A+B*+C*+AB+A <sup>2</sup> *	0.92	0.83	0.90	0.84	0.335	29.47
6	A+B*+C*+AB+BC+A <sup>2</sup> *	0.93	0.82	0.90	0.82	0.337	33.18
7	A+B*+C*+AB+BC+A <sup>2</sup> *+B <sup>2</sup>	0.93	0.81	0.90	0.78	0.337	37.22
8	A+B*+C*+AB+AC+BC+A <sup>2</sup> +B <sup>2</sup>	0.93	0.78	0.89	0.72	0.349	43.69

**Table 5-4A: Summary statistics for eight “best” models for AMBR 3 HCP presented in coded factors. The coded factors A, B and C refer to the cell culture variables media osmolality, seeding density and temperature, respectively.**

	AMBR 3 – HCP	STATISTICS					
Model terms	Model	R <sup>2</sup>	Predicted R <sup>2</sup>	Adjusted R <sup>2</sup>	R <sup>2</sup> k fold	RMSE	AIC
1	C*	0.59	0.51	0.57	0.51	218900	636.08
<b>2</b>	<b>A+C*</b>	<b>0.66</b>	<b>0.55</b>	<b>0.62</b>	<b>0.57</b>	<b>206281</b>	<b>635.18</b>
3	A+C*+AC	0.68	0.53	0.62	0.57	205586	637.15
4	A+C*+AC+A <sup>2</sup>	0.68	0.48	0.61	0.53	210049	640.62
5	A+B+C*+BC+A <sup>2</sup>	0.71	0.49	0.62	0.52	205719	642.56
6	A+B+C*+AB+BC+A <sup>2</sup>	0.72	0.46	0.62	0.47	206660	646.20
7	A+B+C*+AB+AC+BC+A <sup>2</sup>	0.73	0.42	0.61	0.47	209678	650.94
8	A+B+C*+AB+AC+BC+A <sup>2</sup> +B <sup>2</sup>	0.74	0.27	0.59	0.34	215761	657.16